Real Macroeconomic Theory

March, 2014

Per Krusell

Contents

1 Aim **9**

2 A framework for macroeconomics **11**
2.1 From the accounting identity to a resource constraint 12
2.1.1 The aggregate production function and its inputs 12
2.2 Intertemporal macroeconomics 15
2.3 Supply- vs. demand-driven output 15
2.4 Accounting for the growth facts 17
2.5 The time horizon and the population 19
2.5.1 Time 19
2.5.2 People 20
2.5.3 Preferences 22
2.5.4 Choice 22
2.6 Preview of the whole text 24

3 The Solow model **27**
3.1 The model 27
3.2 Applications 30
3.2.1 Growth 30
3.2.2 Business cycles 35
3.2.3 Using the Solow model for other applied macroeconomic questions . . 41

4 Dynamic optimization **43**
4.1 Sequential methods 43
4.1.1 A finite horizon 43
4.1.2 Infinite horizon 50
4.2 Dynamic programming 58
4.3 The functional Euler equation 67
4.4 References 69

5 Competitive equilibrium in dynamic models **71**
5.1 Sequential competitive equilibrium 73
5.1.1 An endowment economy with date-0 trade 73

5.1.2 The same endowment economy with sequential trade . . . 75
5.1.3 The neoclassical growth model with date-0 trade . . . 76
5.1.4 The neoclassical growth model with sequential trade . . . 79
5.2 Recursive competitive equilibrium . . . 80
5.2.1 The neoclassical growth model . . . 81
5.2.2 The endowment economy with one agent . . . 84
5.2.3 An endowment economy with two agents . . . 85
5.2.4 Neoclassical production again, with capital accumulation by firms . . . 86

6 Uncertainty 89
6.1 Examples of common stochastic processes in macroeconomics . . . 89
6.1.1 Markov chains . . . 89
6.1.2 Linear stochastic difference equations . . . 91
6.2 Maximization under uncertainty . . . 93
6.2.1 Stochastic neoclassical growth model . . . 96
6.3 Competitive equilibrium under uncertainty . . . 106
6.3.1 The neoclassical growth model with complete markets . . . 107
6.3.2 General equilibrium under uncertainty: the case of two agent types in a two-period setting . . . 110
6.3.3 General equilibrium under uncertainty: multiple-period model with two agent types . . . 114
6.3.4 Recursive formulation . . . 117
6.4 Appendix: basic concepts in stochastic processes . . . 118

7 Welfare in macroeconomic models 123
7.1 Definitions and notation . . . 123
7.2 An endowment economy . . . 126
7.2.1 Sequential markets . . . 126
7.2.2 Arrow-Debreu date-0 markets . . . 127
7.2.3 Application: endowment economy with one agent per generation . . . 129
7.3 Economies with intertemporal assets . . . 136
7.3.1 Economies with fiat money . . . 136
7.3.2 Economies with real assets . . . 140
7.3.3 A tree economy . . . 140
7.3.4 Storage economy . . . 142
7.3.5 Neoclassical growth model . . . 143
7.4 Dynamic efficiency in models with multiple agents . . . 145
7.5 The Second Welfare Theorem in dynastic settings . . . 148
7.5.1 The second welfare theorem in a 1-agent economy . . . 149
7.5.2 The second welfare theorem in a 2-agent economy . . . 151
7.6 Uncertainty . . . 154
7.7 Hybrids . . . 155
7.7.1 The benchmark perpetual-youth model . . . 156

7.7.2 Introducing a life cycle . . . 158

8 Consumption and labor choice **159**
8.1 A static model . . . 159
8.2 A dynamic model . . . 162
8.2.1 Preference requirements for exact balanced growth . . . 162
8.2.2 Labor-leisure and consumption choice over time without saving . . . 164
8.2.3 A linear savings technology . . . 165
8.2.4 Decreasing marginal returns to savings . . . 167

9 Growth **169**
9.1 Some motivating long-run facts in macroeconomic data . . . 169
9.1.1 Kaldor's stylized facts . . . 169
9.1.2 Other facts . . . 170
9.2 Growth theory I: optimal steady states and convergence . . . 171
9.2.1 Properties of the capital accumulation function . . . 172
9.2.2 Global convergence . . . 176
9.2.3 Dynamics: the speed of convergence . . . 178
9.3 Growth theory II: exogenous technological change . . . 185
9.3.1 Exogenous long-run growth . . . 186
9.3.2 Choosing to grow . . . 191
9.3.3 Transforming the model . . . 192
9.3.4 Adjustment costs and multisector growth models . . . 194
9.4 Growth theory III: endogenous growth . . . 195
9.4.1 The AK model . . . 195
9.4.2 Romer's externality model . . . 198
9.4.3 Human capital accumulation . . . 200
9.4.4 Endogenous technological change . . . 201
9.4.5 Directed technological change . . . 206
9.4.6 Models without scale effects . . . 206
9.5 What explains long-run growth and the world income distribution? . . . 206
9.5.1 Long-run U.S. growth . . . 206
9.5.2 Assessing different models . . . 206
9.5.3 Productivity accounting . . . 208
9.5.4 A stylized model of development . . . 208

10 Business cycles **211**
10.1 Introduction . . . 211
10.2 Stylized facts . . . 213
10.2.1 Early Facts . . . 213
10.2.2 Kydland and Prescott (1982): How to convert raw data into facts . . . 214
10.2.3 Kydland and Prescott's facts . . . 215
10.3 Real business cycle theory: the basic model . . . 217

10.3.1 Basic methodology: Introduction to calibration . . . 217
10.3.2 Measurement: Solow growth accounting . . . 217
10.3.3 Real business cycle models: brief cookbook procedure . . . 220
10.3.4 Real business cycle in action: A model example . . . 225
10.4 Other real business cycle models . . . 226
10.4.1 Investment-specific technology shocks . . . 226
10.4.2 Fiscal policy and government spending shocks . . . 226
10.4.3 Monopolistic competition and markup shocks . . . 226
10.4.4 Preference shocks . . . 227
10.4.5 Models with adjustment costs in technology and preferences . . . 227
10.5 Sunspot models of the business cycle . . . 227
10.6 References . . . 227

11 Frictional labor markets 229
11.1 Taxes in a labor market without frictions . . . 230
11.1.1 Proportional taxes on labor earnings . . . 230
11.1.2 Data . . . 231
11.2 Indivisible labor . . . 231
11.3 Theories of unemployment . . . 233
11.3.1 Search theory . . . 233
11.3.2 The Pissarides model . . . 235
11.3.3 The European unemployment dilemma . . . 240
11.3.4 Rogerson . . . 240
11.4 Inequality in the labor market . . . 240
11.4.1 Changes in K . . . 241

12 Asset pricing 243
12.1 Lucas' Asset Pricing Formula . . . 243
12.1.1 The Model . . . 243
12.1.2 Solving for prices of state-contingent claims . . . 244
12.1.3 Pricing assets . . . 244
12.2 The Equity Premium Puzzle . . . 246
12.2.1 The Model . . . 247
12.2.2 Calibration . . . 251
12.3 Suggested Solutions to the Puzzle . . . 251
12.4 References . . . 254

13 Economic policy 257
13.1 Ricardian equivalence . . . 258
13.1.1 Dynastic model . . . 258
13.1.2 Overlapping generations . . . 260
13.2 Optimal distortionary taxes . . . 263

13.2.1 Taxation smoothing . . . 263
13.2.2 Optimal taxation mix . . . 266
13.3 References . . . 272

14 Aggregation **273**
14.1 Inelastic labor supply . . . 273
14.2 Valued leisure . . . 276
14.2.1 Wealth effects on labor supply . . . 276
14.2.2 Wealth effects on labor supply . . . 277

Chapter 1

Aim

These lecture notes cover a one-semester course in macroeconomics. The aim is to provide a basis for learning the main methodological tools in modern macroeconomics while at the same time providing a survey of the main questions and answers given in the modern literature. The emphasis is on "quantitative theory", i.e., theory specified so as to match basic features of the data and used for addressing quantitative questions. This means that much of the notes will be focused on developing a theoretical toolbox. The notes will accomplish this mostly by applying rather standard microeconomic notions to a macroeconomic context. Thus, a solid understanding of applied microeconomic theory is a very useful background, if not a prerequisite. As for statistical methods, the main quantitative applications will not be based on formal econometrics. Therefore, the presumption is that students invest separately in the econometric techniques needed for formal model estimation, classical or Bayesian. Likewise, and as the title of the text suggests, the models here are "real", i.e., they do not analyze monetary issues. Thus, the student is expected to consult another text for this material; at the same time, the maintained belief here is that modern monetary theory requires a thorough course in real macreoconomics so it would make sense to read the present text, or one like it, first. In fact, the whole idea of this text is not to cover everything but rather provide a reasonably thorough treatment of material that really is hard to skip: all of the material should be included in any first-year graduate course in macroeconomics. At times, as an author I occasionally feel somewhat embarrassed that the manuscript is not more advanced than it is, and then I try to remind myself that the basics still need to be covered, and I am not sure that there is another text that accomplishes this job.

Chapter 2 briefly summarizes the main long-run facts for the key macroeconomic variables and outlines a framework that can be used to account for them both qualitatively and quantitatively. The concluding part of Chapter 2 also provides a roadmap for the rest of the course, which essentially studies the macroeconomic framework at greater depth and in a variety of applications. Two blocks of chapters thus follow: Chapters 3–7 cover methods—illustrations of how the main framework can be used, optimization, equilibrium definitions, uncertainty, and welfare results—and the remaining chapters cover applications, beginning with growth and business cycles.

Chapter 2

A framework for macroeconomics

The purpose of this chapter is to propose a broad theoretical framework for addressing macroeconomic questions. The framework will strive to build as much as possible on microeconomic foundations, and hence much of the text will use standard microeconomic tools. Occasionally, the approach using explicit microeconomic foundations will be described as "modern macroeconomics", a term which is sometimes used in the literature. Almost all macroeconomic models used in research today are of this kind. Their history, however, is not that long: the first models began to be formulated in the 1970s. In constructing the macroeconomic framework, the guiding principle will be an aim to organize the main macroeconomic facts: the aim is to construct a "quantitative theory". It is also important that the framework be broad enough to encompass the main areas of study, including growth and business cycles and the many subtopics of these areas usually identified in graduate macroeconomic classes.

As for the main facts, the variables are the main aggregate quantities and prices: output, consumption, investment, the capital stock, labor input/hours worked, wage rates, real interest rates, unemployment, and some more. All the facts discussed here will be described in a rather stylized way; for example, the rate of unemployment will be described to be "stationary", which many economists would argue against, since we have witnessed rather persistent and quantitatively large increases in unemployment on several occasions (in the 1970s in Europe and recently more broadly). Here, "stationary" thus does not preclude important and persistent swings, but rather should be interpreted as saying that there does not appear to be a consistent drift toward 0 or 1. The main facts we emphasize here are long-run facts; short-run facts are discussed in more detail later. The growth facts we will focus on are as follows [AT LEAST ONE GRAPH ON EACH, U.S. DATA]:

1. output per capita has grown at a roughly constant rate

2. the capital-output ratio (where capital is measured using the perpetual inventory method based on past consumption foregone) has remained roughly constant

3. the capital-labor ratio has grown at a roughly constant rate equal to the growth rate of output

4. the wage rate has grown at a roughly constant rate equal to the growth rate of output
5. the real interest rate has been stationary and, during long periods, roughly constant
6. labor income as a share of output has remained roughly constant
7. hours worked per capita have been roughly constant.

2.1 From the accounting identity to a resource constraint

The basic accounting identity in undergraduate textbooks is $C + I + G + NX = Y$: consumption plus investment plus government expenditures plus net exports equals GDP. The next steps of our analysis will be a structural reformulation of this identity. We will begin by abstracting from government and foreign trade: we consider a simple "closed" economy where $G = NX = 0$ at all times. Second, we will interpret the identity as a resource constraint: $C + I = Y$ then means that Y is an amount of resources available and it can be spent on either C or I, or any combination of the two (so we can have $C = Y$ or $I = Y$ as extreme cases). This means, in effect, that we think of consumption and investment as perfect substitutes. Of course they are not, at least not in the short run—although some goods can be used literally as both consumption and investment (a range of durable goods), most cannot, and it takes take to reorganize society's production from one type of good to another. Third and last, we will specify where resources originate from by appeal to an aggregate production function: a function that specifies the total amount of resources available as a function of basic production inputs: capital and labor. I.e., we will write $Y = F(K, L)$, so that we have a resource constraint that reads $C + I = F(K, L)$.

2.1.1 The aggregate production function and its inputs

The aggregate production function plays an important role in much of macroeconomic theory. It is a simple representation of how output is produced from the basic inputs of an economy. We will use a set of standard assumptions about the aggregate production function throughout the text, and the present section will list them and discuss them briefly. Some of this discussion will be microeconomic in nature, but it is important as so much of macroeconomic analysis relies on these assumptions.

We most often assume that the aggregate production function (i) has constant returns to scale (CRS: F is homogeneous of degree one in its input vector); (ii) is strictly increasing in each input; (iii) is strictly quasiconcave, i.e., such that the isoquants are convex toward the origin.[1] CRS is motivated, briefly, first by ruling out decreasing returns to scale by a replication argument: if a certain output level can be produced from given inputs, it should

[1]Equivalently, $F(K, L)$ equal to a constant implicitly defines K as a function of L that is convex, no matter what the constant is.

be possible to double all inputs and obtain at least double output by just doing the same thing all over once more. Increasing returns, on the other hand, seem plausible from a variety of perspectives and for various specific production processes, but here the maintained view is that the extent of increasing returns seems small enough to be well approximated by zero. This argument is usually made with reference to the empirical literature on production function estimation, where large departures from constant returns are rarely recorded even in disaggregated data. As we shall see in different sections of the text, departures from constant returns are actually not uncommon in macroeconomic modeling. Departures from strict quasiconcavity are used as well, but then more as illustration or for the purposes of easy model solution.

Having stated the basic assumptions about F, several points are important to note here. First, whereas capital and labor are the main inputs for most firms, typical firms also have many other inputs: intermediary goods and services purchased by other firms. For example, a restaurant buys food ingredients from suppliers, uses financial-sector services, buys insurance, and so on. However, intermediate inputs, which are inputs of some firms but, by definition, outputs of other firms, are not included in the aggregate production function. The accounting identity we started from, as well as the aggregate production function, can also be viewed as the total "value added" of the firms in the economy: the sum of what each firm adds over and beyond the value of the intermediate inputs it uses.

Second, what about basic raw materials? Refined raw materials, such as gasoline used as vehicle fuel, are intermediate inputs, so they should not appear as production inputs. However, "literally raw" materials should perhaps be included: the amount of petroleum (unprocessed oil) used up in production, to the extent it has market value, arguably should, as should basic other raw materials in their crude form (metal ore, etc.). Some basic inputs, such as trees used to produce timber, are really produced—in a process that takes years—using labor, capital, and intermediate inputs, and should hence not be included. Iron ore could be thought of the same way but because the formation of iron ore takes thousands of years and is not even well understood it is not commercially viable and therefore is regarded as a non-renewable resource, as opposed to timber trees which are renewable.[2] In sum, in some of our analysis it can be important to use a third basic input, $F(K, L, M)$, where M then stands for non-renewable materials. In this case, it would also be important to keep track of the remaining stock of M at all points in time: a capital stock of sorts. However, in almost all macroeconomic studies, basic raw materials are abstracted from since they account for a very small share of costs among the aggregate of firms (compared to the capital and labor costs).

Third, one can imagine expressing a resource constraint where C and I are produced from the basic inputs K and L in a more general way. So consider a general formulation $G(C, I, K, L) = 0$ for the production possibility frontier. This formulation allows C and I to be less than perfect substitutes, and it moreover allows the tradeoff between C and

[2]Formally, also air, water, and other basic inputs are used in production, but to the extent they are available for free and in "unlimited supply", they are not included as inputs. In many places on earth, some of these inputs, such as water in many African economies, is scarce and, at least in principle, should be treated as an intermediary input.

I to depend on the inputs available. One can alternatively consider the production of C and I separately, in two sectors, each with its own production technology: $C = F_c(K_c, L_c)$ and $I = F_i(K_i, L_i)$, where $K_c + K_i = K$ and $L_c + L_i = L$. This intuitive formulation amounts to a form of G with the above specification. To see this, suppose we aim for a given $C = \bar{C}$. We can then formulate the largest I possible—that on the production possibility frontier—as a maximization problem: I equals the maximum, by choice of K_c and L_c, of $F_i(K - K_c, L - L_c)$ under the restriction $\bar{C} = F_c(K_c, L_c)$. This maximization is well defined given regularity assumptions on the two sectoral production functions and defines, for each $\bar{C}$ and values for K and L, an I, hence mapping out a G.[3] For many analyses it makes sense to go beyond the simple $C + I = F(K, L)$ formulation, and the present text will give a number of such examples.

Especially when we consider the macroeconomy evolving over shorter periods of time—as we will on and off throughout the book—it becomes important to think about how "flexible" it is in terms of moving resources between the C and I production uses, and the just mentioned generalization of the resource constraint, $G(C, I, K, L) = 0$, allows one to consider less flexibility than the perfect-substitutability setting given by our benchmark $C+I = F(K, L)$. However, a perhaps more intuitive way to think about inertia is to have the capital stock be pre-committed to sectors, or at least harder to move across sectors. The pre-commitment case would thus call for a multidimensional (by sector) capital stock—a vector (K_c, K_i)—such that at any point in time the capital stocks are committed to, and immobile from, the sector, but over time new investments can change the total future stocks in each sector. Multidimensional capital stocks are considered in some of the macroeconomic literature and will be briefly discussed below; it primary use is to capture how the economy is less than fully flexible in responding to various forms of shocks calling for sector reallocation. Similar arguments can be made for the labor force, which may need to be retrained in order to move across sectors.

Finally, the aggregate nature of the production function used in most of macroeconomic theory calls for a defense; in reality many, many different kinds of goods and services are produced, and in many locations. Put differently, when can production of different goods and services be aggregated into one function? There are theoretical conditions under which precise aggregation is possible. Consider, for example, two goods and two production functions; then if inputs are fully mobile, if the production functions have identically-shaped isoquants, there is an aggregate production function representation.[4] It is easy to come up with examples that would violate the assumption, on the other hand, although no systematic evidence has been gathered to suggest that departures from the assumption are quantitatively important in practice. It is, however, still an open question what the link is between plant-level production technologies and the kinds of aggregate technologies for GDP generation macroeconomists use in practice; it may, for example, be that aggregation is a good

[3]Can this alternative formulation alternatively be described as the more restrictive $G(C, I) = F(K, L)$?

[4]For two goods x and y, with $x = A_x F(k_x, l_x)$ and $y = A_y F(k_y, l_y)$, where A_x and A_y can differ, $k_x + k_y = k$ and $l_x + l_y = l$ implies an aggregate representation $x + \frac{A_x}{A_y} y = A_x(k, l)$, where $\frac{A_x}{A_y}$ is also the market relative price of y.

approximation under some but not all circumstances (such as during economic crises).

2.2 Intertemporal macroeconomics

Can the facts above—that concern long-run features of the data—be made consistent with the resource restriction we just introduced? According to our framework, output growth must come from growth in inputs or growth in the output produced by given inputs—continuing upward shifts in the production function F. (We will think of the resource constraint in per-capita form so growth in population will not be helpful for raising output per capita.) Thus, we attach time subscripts and write $C_t + I_t = F_t(K_t, L_t)$, t representing a time period, which in different applications will have a different length. We will link time periods as follows. For capital, we assume $K_{t+1} = (1 - \delta)K_t + I_t$, where δ is the rate of depreciation of capital, which is assumed constant here (but could of course be thought to vary or even be endogenous, as it is in many studies). Thus, investment formally builds future capital.

As for the labor input, two features are important to consider. One is the idea that labor can be varied by choice—in response to changes in the economic environment, households may choose to supply more or less labor to the market. One can of course also imagine that unemployment prevents the whole labor force from being employed in production, a case we will of course discuss at length below and which will commented further on momentarily. The other is that the "quality" of the labor input can change over time: human capital accumulation. Human capital accumulation is a key element of long-run macroeconomic analyses but is typically abstracted from in business-cycle analysis.

Also the production function is allowed a time subscript, F_t, indicating that the production possibilities from given inputs can improve over time: technological change. Just how the production function changes over time is a whole topic in itself and it will be discussed at length below in the context of growth analysis.

Finally, one can imagine—in the version of the model where material resources are explicitly considered—growth due to an increasing use of raw materials. Indeed, many people—especially non-economists perhaps—view increasing raw material use as a key behind why our economies have been allowed to grow over time, so our analysis of economic growth below will have to discuss this possibility. For our benchmark, and indeed in the vast majority of traditional macroeconomic models (micro-founded or not), raw material use is abstracted from both in the short and in the long run.

2.3 Supply- vs. demand-driven output

Before moving on, let us note that the resource constraint $C + I = F(K, L)$, together with a capital accumulation equation and assumptions about how F and L evolve over time, could be labeled a "neoclassical" representation of key macroeconomic aggregates. The reason is that it presumes that output is a direct function of all the inputs available: output is driven by "supply". Thus, the equation does not in an obvious way admit a "demand determination of output", since output is simply given by a technological specification and the basic inputs available for production. Demand determination would, for example, mean that an increased willingness to consume—a higher C—could be accommodated without a decrease in investment—without lowering I—by simply raising Y appropriately. This could not really occur by an increase in the amount of capital, since the capital stock—machines, buildings, etc.—is given from past investment decisions and cannot be increased in the short run. Labor input, of course, could: how many hours are supplied to the market is fundamentally endogenous and might respond to consumption demand; one can more generally imagine that labor effort is determined by a number of factors, including consumer's expectations of future economic conditions. To admit a fuller representation of demand effects once could incorporate two additional variables: the utilization rates of both the capital stock and the labor force. Then, the resources could be written $F(u_k K, eL)$, where u_k is the utilization rate of the available capital stock and e is the employment rate ($1-e \equiv u$ is the unemployment rate). Now, clearly, u_k and e could be perceived to depend on demand factors, though of course a precise form of these relationships would need significant further theorizing.

We will return to how the modeling of variable utilization of inputs could be accomplished later in the text. Over short horizons, indeed, utilization rates are broadly believed to be variable and can hence be an important element of understanding how aggregate output varies over time. Current business-cycle models—at least those used by policymakers in central banks and elsewhere—certainly take them into account, though it is fair to say that most of the many early versions of the modern macroeconomic models (say, from the 1980s and early 1990s), abstract from utilization variation over time.

For long-run analysis, however, it makes sense to abstract from variations in utilization. This abstraction relies primarily on the available data: it does not appear that either the utilization of capital or the unemployment rate display long-run trends. Indeed, it would also be surprising from a theoretical perspective if they featured persistent change in one direction. One could imagine, for example, that an economy would permanently move from an average low rate of unemployment to a higher one, say, if certain forms of institutional change occurred in the labor markets (such as through regulatory changes). Permanently ongoing change, with ever-increasing or ever-decreasing rates, would suggest that there is no bound to institutional change, which does not appear plausible for modern economies.

Notice that the last argument implies, rather directly, that demand could not by itself be driving economic growth in the longer run. That is, economies do not grow because households have ever-increasing demand for goods. Of course, consumption growth will require demand to expand, but the key here is that it will not be caused by demand growth. Hence, statements like "we need to boost demand to induce economic growth", common in policy discussions, might well make sense but only when "growth" is interpreted as a short-run phenomenon. Similarly, the idea that monetary policy would influence long-run growth can be dispensed with at the outset. For these reasons, to start with, because the initial focus will be on long-run growth, we will assume full utilization of both the capital and labor inputs. (One could alternatively simply assume a constant utilization rate—any number between 0 and 1—of both factors, but to save on notation let us use $u_k = e = 1$.)

2.4 Accounting for the growth facts

First of all, we will think of long-run growth in terms of the concept *balanced growth*. Growth is balanced when all variables grow at exactly constant rates. These rates can, and will, be different for different variables; in particular, along a balanced growth path, some variables will remain constant. Exactly constant growth rates may be too strict a requirement—one can imagine asymptotic, though not exactly constant, growth rates—but we will nevertheless focus on this case mostly; a key reason is that it is analytically easier. But to insist on approximately constant rates is important, since modern economies have satisfied the growth facts over a very long period of time.

To satisfy the first fact, we need to be specific on the source of growth. We will simplify the analysis here by assuming that there is no long-run human capital growth but that there is technological growth of a specific kind: it is labor-augmenting, i.e., it is equivalent to raising labor input over time. Thus, we consider a production function $F_t(K, L) = F(K, A_t L)$, where A_t is a time series that can display growth. We will formulate growth rates in gross terms. Thus, we will focus on the case $A_t = \gamma^t$, thus normalizing so that $A_0 = 1$ and assuming labor-augmenting growth at rate γ. Notice now that for a linear function of several variables—a simple case of which is $C_t + I_t$—to display (i) constant growth while (ii) the individual variables all grow at constant rates, we necessarily need all the variables to grow at the same rate. This means, in our particular context, from the assumption that $C_t + I_t$ will grow at a constant rate and $K_{t+1} = (1 - \delta)K_t + I_t$, that C, I, and K must all grow at the same rate: the rate of growth of output. Furthermore, let us use the assumption for now that labor input is fixed over time; this assumption is consistent with fact 7 listed above, the constancy of labor input, and of course it remains to be accounted for on a deeper level. We will indeed return to this issue shortly. So since the aggregate production function has constant returns to scale, $F(K_t, \gamma^t L) = \gamma^t F(\frac{K_t}{\gamma^t}, L)$, which tells us that output growth equal to γ is feasible: if output grows at γ, so must K_t, and hence both arguments of F in the previous expression are constant, verifying that output indeed grows at rate γ. To summarize this discussion formally, suppose that $K_t = \gamma^t K_0$, $I_t = \gamma^t I_0$, and $Y_t = \gamma^t Y_0$; then we see that the resource constraint is satisfied at all times so long as $C_0 + I_0 = F(K_0, L)$ and that the capital accumulation equation is satisfied at all times so long as $\gamma K_0 = (1 - \delta)K_0 + I + 0$. This gives us two restrictions on the initial values of all variables: $I_0 = (\gamma + \delta - 1)K_0$ and $C_0 = F(K_0, L) + (\gamma + \delta - 1)K_0$. Thus, for each value of K_0 there is an associated investment, consumption (and output) level, and so different balanced growth paths can be followed—but situated on different levels.

The previous discussion suggests that it is feasible to grow so as to satisfy facts 1–3 above. However, for any given initial capital stock, exact balanced growth would require a specific value of investment (or, equivalently consumption); is it reasonable to expect this value to be chosen in the data? This is where Solow's growth model comes in: Solow (1956) turned the question around and rather argued that it was reasonable to assume that modern economies have a constant rate of saving, and he then showed that a constant saving rate will be asymptotically consistent with balanced growth for any initial capital stock (and exactly consistent with balanced growth for exactly one level of the initial capital stock).[5] Thus, Solow assumed that for some fixed saving rate s, $I_t = sY_t$ for all t. Thus, we can write the evolution of the capital stock in the economy as

$$K_{t+1} = sF(K_t, \gamma^t L) + (1-\delta)K_t$$

for all t. We can use a simple variable transformation now—let $\hat{K}_t = K_t/\gamma^t$—to obtain

$$\gamma \hat{K}_{t+1} = sF(\hat{K}_t, L) + (1-\delta)\hat{K}_t.$$

Thus, the equation defines a (nonlinear) first-order difference equation in transformed capital. Solow showed, as will we in Section **??** below, that under some suitable, and reasonable, assumptions on the primitives—γ, s, F, and δ—the dynamics imply global, and monotone, convergence to a *steady state* for $\hat{K}_t$, i.e., $\hat{K}_t$ is a monotone sequence and there exists a $\hat{K}_{ss} > 0$ such that $\lim_{t\to\infty} \hat{K}_t = \hat{K}_{ss}$ for all $K_0 > 0$. The steady state is thus given by the solution to $(\gamma + \delta - 1)\hat{K}_{ss} = sF(\hat{K}_{ss}, L)$. The key point now is that as $\hat{K}_t$ converges to its steady state, K_t moves closer to its balanced growth path. It is also straightforward to verify that, at the same time, Y_t moves toward its balanced growth path. Thus, the convergence result allows us to verify that for any given s—our behavioral assumption at this point—the economy will move towards a path satisfying facts 1–3.

To address facts 4–6 we need to talk about market allocations and prices. We will think of firms as operating in perfect competition, which with a constant-returns technology is consistent with zero profits in equilibrium. It is easiest to think of firms as having static profit maximization problems: $\max_{k,l} F(k, A_t l) - r_t k - w_t l$, where we use the output good as the numéraire, r_t is the rental price for capital—the price paid to capital owners for using it in period t—and w_t is the wage rate. In an equilibrium, both k and l will be positive, so the first-order conditions, evaluated at equilibrium quantities, will read $r_t = F_1(K_t, A_t L)$ and $w_t = A_t F_2(K_t, A_t L)$. Because F has constant returns to scale—is homogeneous of degree one in the input vector—the partial derivatives are homogeneous of degree 0, so we have $r_t = F_1(\frac{K_t}{A_t L}, 1)$ and $w_t = A_t F_2(\frac{K_t}{A_t L}, 1)$. Given that K_t and A_t grow at the same rate we conclude that, on a balanced growth path, r_t is constant and w_t grows at the same rate of output. The latter observation is fact 4. The former is related to fact 5: the return on a unit of investment, in a competitive economy, will equal the real interest rate. A unit of investment costs one output unit at t and delivers, at $t+1$, a return r_{t+1} (the productive use of the capital next period) plus $1-\delta$, since the capital does not disappear but only depreciate a bit. Thus, with δ constant, the gross interest rate is thus constant over time whenever r_t is constant over time: fact 5.

Finally, fact 6 states that the labor income as a share of total income has stayed roughly constant over time. The labor income share is $w_t L/F(K_t, A_t L)$, which in turn equals $A_t L F_2(\frac{K_t}{A_t L}, 1)/(A_t L F(\frac{K_t}{A_t L}, 1)) = F_2(\frac{K_t}{A_t L}, 1)/F(\frac{K_t}{A_t L}, 1)$. (Notice that, in terms of the theory

[5] Swan (1956) provides the same result.

adopted here, the capital share, $r_t K/F(K_t, A_t L)$, must equal 1 minus the labor share, since F has constant returns to scale, so that there are no profits.[6]) Clearly, this object is constant along the balanced growth path. However, since the labor share really has displayed very minor fluctuations, macroeconomists usually restrict F to a specific function that literally delivers a share that is constant at all times. A function which satisfies this property—indeed the only such function—is the Cobb-Douglas function: $F(k,l) = Bk^{\alpha}l^{1-\alpha}$.[7]

It remains to provide a fuller explanation of why, over a long period of time (i) saving rates are approximately constant (which allowed us to match several of the facts) and (ii) the amount of hours worked is approximately constant (fact 7). For this, we need to be explicit about household choice: saving rates are determined by people, as are working hours. The next section thus discusses different population structures commonly used in macroeconomic models and then demonstrates how certain assumptions on the population of households allows us to rationalize these facts.

Before concluding, it must be pointed out that the details of the macroeconomic setting developed above, a setting whose primary purpose is to accounting for the salient long-run macroeconomic facts of developed economies, are not merely sufficient for generating the facts but also necessary. For example, we assumed that technological change is labor-augmenting in nature. In the chapter on growth below we will show that, unless the production function is Cobb-Douglas, no other form of technology growth is consistent with balanced growth.[8] The preference specifications below, which generate constant saving rates and constant labor hours, are also uniquely pinned down.

2.5 The time horizon and the population

We discuss the topics of time, people, and people's preferences first. We then argue that certain restrictions on preferences will be consistent with the growth facts: constant saving rates and constant hours worked.

2.5.1 Time

In practice, the intertemporal macroeconomic framework used in research almost always involves an infinite number of time periods (we mostly use discrete, and not continuous, time in this text). The reason for this is mainly analytical convenience: it allows stationarity in

[6]Formally, Euler's theorem states that for a function F that is homogeneous of degree 1, $kF_k(k,l) + lF_l(k,l) = F(k,l)$ for all (k,l).

[7]A heuristic way of proving this is as follows. For simplicity, look at the capital share instead, i.e., $K_t F_1(\frac{K_t}{A_t L}, 1)/(A_t L F(\frac{K_t}{A_t L}, 1))$. Suppose we define $f(x) \equiv F(x,1)$; then, clearly, $F_1(K,1) = f'(x)$, and the constancy of the share implies that for some share value α, $xf'(x)/f(x) = \alpha$ for all x. Thus we have that $f'(x)/f(x) = \alpha/x$. We can write this as $d\log f(x)/dx = \alpha d\log(\hat{B}x)/dx$ for all x and any $\hat{B}$. Rewriting, we obtain $d\log f(x)/dx = d\alpha\log(Bx)/dx = d\log(\hat{B}x)^{\alpha}/dx$. Thus, since this equation has to hold for all x, and defining $B = \hat{B}^{\alpha}$, we have that $f(x) = Bx^{\alpha}$. This means that, with $x = K/(AL)$, we have $F(K, AL) = ALF(K/(AL), 1) = ALB(K/(AL)^{\alpha} = BK^{\alpha}(AL)^{1-\alpha}$.

[8]The proof uses exactly the same technique as in footnote 9 below.

the sense that the remaining time horizon is the same—it is infinite—as time passes. In a model with a finite number of periods, this is not true, but sometimes finite-horizon models are used as well. Finite-horizon models, such as a two-period model, can be very useful for understanding intertemporal macroeconomic mechanisms, because they may allow more analytical clarity. In quantitative theory, however, it is key that the time horizon be long enough that the idea that "the world ends soon" not influence the results too much, such as it would, for example, in a two-period model. Thus, for now we will assume that the time horizon is infinite: $t = 0, 1, \ldots$.

2.5.2 People

Regarding the description of households, there are several important aspects that need to be considered in choosing an analytical framework. First, the most commonly used assumption—and the one we will use here—is that households are "unitary", that is, treated as one individual, even though in practice most households have more than one member. It is not altogether uncommon in modern macroeconomic models to see households modeled as consisting of two adult members, in which case one needs to also be clear on how joint decisions are made, and here there are several paths followed in this literature. Relaxations of the unitary model can be important for understanding how changing patterns in marriage dynamics—the prevalence of marriage, divorce, and associated regulations—may have influenced the macroeconomy through aggregate saving, risk taking, or labor supply. The simplicity of the unitary model is our main reason for adopting it here, but it should also be pointed out that for most of the applied topics in this text, a more general treatment would arguably lead to very similar analysis and conclusions.

Second, one must specify for how many periods a household lives. Although it might seem obvious that the "right" assumption is that of a finite life—one that in quantitative analysis would be calibrated to match average lifetimes, perhaps allowing for the uncertainty of death—in practice the most common assumption is that households live forever. This, however, is not unrealistic if we think of households as dynasties, where current members derive utility both from their own consumption and that of their children, grandchildren, and so on. However, altruism has to be "perfect": it has to be that the current decision maker shares the preferences of the future dynasty members. It is easy to imagine cases where this would not be true. One case may be particularly pertinent: suppose that the current decision makes places a much higher weight on him- or herself than on any descendant, and that the weights on future descendants are rather similar. In this case, the current decision maker would like to treat future generations similarly, but any future decision maker would disagree with that and put a higher weight on him- or herself. Such a setting, thus, involves *time inconsistency* of the dynastic preferences, and household choice would formally need to be treated as a game between different players, indexed by time and endowed with (conflicting) preferences. This case has received some attention in the recent macroeconomic literature but requires considerably more complex analytical methods: dynamic game theory, instead of just the single optimization problem that applies for the dynastic, infinitely-lived household.

An alternative, and rather commonly used setting involves "overlapping generations". In its simplest case, suppose that each household lives for two periods and that they have no offspring (or have offspring but have no altruism toward them). Then at any point in time the economy is inhibited by "young households" and by "old households". Such a population would still involve a saving decision—that of the young—but this saving decision is much easier to solve analytically than the infinite-horizon optimization of the dynasty (where the number of choice variables is infinite). However, the overlapping-generations economy has more complicated equilibrium interactions, as we shall see in future chapters, and hence on a purely analytical level, there are both pros and cons with each setting. When used for quantitative analysis, overlapping-generations models have households that live for many periods (where a period would be a year). There are also models with random life lengths; we will also see examples of such models in the text.

Fourth, just like for the aggregate production function it is important to note that in the real world there are many dimensions of heterogeneity that probably ought to be addressed in macroeconomic models relying on microfoundations. Real-world households do not only differ in age, but in a number of in ways: different households have different labor incomes, different levels of wealth, and seem to display different preferences, such as across different kinds of goods and services, between consumption and leisure, and in terms of attitudes toward risk. However, most macroeconomic models focus on the homogeneous case: they assume a *representative agent*. This assumption is chiefly made for the sake of analytical tractability, and it will be maintained in most of this text. As with aggregation across firms, there are also assumptions—regarding consumers' preferences, along with a market structure—that allow significant differences across households and yet allow aggregation, i.e., allow a single-household representation of the collection of all the households (we will demonstrate and discuss this in later chapters). However, these assumptions are not all realistic, and the study of the relevance of consumer heterogeneity for macroeconomics is an important and active ongoing research area. In some respect this area is rather mature; there are many macroeconomic models with consumer heterogeneity and there is also a growing empirical literature where microeconomic data is used to assess the extent and importance of heterogeneity. A tentative summary of the findings in this area is that consumer heterogeneity seems to be very important for some macroeconomic questions but that there are many issues for which the representative-household assumption would appear adequate. Thus, the representative-agent assumption appears to be a very useful benchmark at the very least.

Finally, we will assume that households are fully rational, i.e., that their preferences are represented by a utility function that they maximize when they make their choices. Perfect foresight and rational expectations (the applicable term in the case where consumers face uncertainty) are then simply one aspect of utility maximization. The rationality assumption is of course extreme but remains a useful, and very highly utilized, benchmark. The rapidly growing literature on behavioral economics certainly has representation in macroeconomics as well but so far no "new standard" has received broad enough support as an alternative to the assumption of unlimited rationality.

2.5.3 Preferences

Households will be assumed to have utility functions that are time-additive, with consumption in periods t and $t+1$ evaluated as $u(c_t) + \beta u(c_{t+1})$, where u is strictly increasing and strictly concave. Thus, the same function u is used for consumption in both periods, but there is a weight β on $t+1$ consumption. The fact that u is the same for both consumption goods implies that consumption in both periods are normal goods: with more income, consumers would like to consume more of both goods, which seems very reasonable in this application. Another aspect of this setting is an element of *consumption smoothing*: there is decreasing marginal utility to consumption in both periods so spending all of an income increase in one period will in general never be optimal. Furthermore, $\beta < 1$ captures impatience, or a probability of death—or any other reason for down-weighting future utils—and will be a typical assumption.

2.5.4 Choice

An important part of the text will explain intertemporal choice from first principles: different methods for solving intertemporal problems, with and without uncertainty, along with a number of important macroeconomic applications. Here, the purpose is to very briefly explain the key steps, heuristically, in order to explain the growth facts.

Conceptually, the way consumers make decisions—if able to choose when to consume their income—is according to basic microeconomic principles: so as to set their marginal rate of substitution equal to the relative price. We will now go through the two key choices using these principles.

Consumption vs. saving

The relative price between consumption and t and $t+1$ is the *real interest rate*: it is the amount of goods at $t+1$ that a consumer can buy for one unit of the good at t. We will denote the gross real interest rate between t and $t+1$ R_{t+1} here. The marginal rate of substitution between the goods can be obtained by defining an indifference curve relating to these two goods, $u(c_t) + \beta u(c_{t+1}) = \bar{u}$ and taking total differentials, i.e., $u'(c_t)dc_t + \beta u'(c_{t+1})dc_{t+1} = 0$ and solving for dc_{t+1}/dc_t. Setting the resulting expression equal to the gross real interest rate, we would thus have

$$\frac{u'(c_t)}{\beta u'(c_{t+1})} = R_{t+1}.$$

This equation, which equivalently can be written $u'(c_t) = \beta u'(c_{t+1})R_{t+1}$, is commonly referred to as the *Euler equation* and it is a central element of macroeconomic theory. It says that an optimizing consumer sets the marginal utility loss, the left-hand side, of saving one consumption unit for tomorrow equal to the gain tomorrow in consumption terms, R_{t+1}, i.e., the return on the savings, times the marginal utility of each unit tomorrow, $\beta u'(c_{t+1})$.

We showed above that a constant saving rate will simply mean a constant level of capital relative to technology, i.e., $\hat{k}_t$ becomes constant—this is implied by Solow's analysis, which we will elaborate more on later. This also means, from the capital accumulation equation, that the investment-to-technology level becomes constant, which in turn implies, from the aggregate resource constraint, that the consumption-to-technology level becomes constant. In other words, a constant saving rate is associated with aggregate consumption growing at a constant rate. We also saw that balanced growth requires a constant real interest rate. So the question now is whether the Euler equation could hold on for constantly growing consumption. The answer is no for most functions u. However, for a specific class of functions it is possible: the power functions, such as $c^{0.5}$ or $-1/c$. To see this, notice that what we need to require on a balanced path is that $u'(c_t)/u'(gc_t)$ be constant, where g is the constant rate of growth. But this is true if u is a power function: then $u'(c_t)/u'(gc_t)$ becomes a constant, namely, g to the negative of the relevant power (in the examples $g^{-0.5}$ and g, respectively). More generally, the result holds if and only if $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$ for $\sigma \geq 0$, which in the special case $\sigma = 0$ becomes $\log c$.[9] It is easy to verify the "if" part: just verify that with the stated utility function, constant growth in c (at a rate that depends on βR) satisfies the Euler equation. The "only if" result is a little harder but is proved in Chapter 8.

The above discussion has been carried out in terms of simply selecting two adjacent time periods, without reference to how many periods the consumer lives in total. This also means that the arguments above hold whether in dynastic or overlapping-generation economies, or some combination of these, so that the restrictions placed on preferences in order to be consistent with balanced growth hold rather generally.

Labor vs. leisure

Turning to labor supply, with a similar level of generality, the period utility function then has to allow for a value of leisure: $u(c, l)$, where l denotes leisure. We thus need to insist on balanced growth in consumption jointly with a constant labor supply in the long run. We therefore need the condition $\frac{u_l(c_t,l_t)}{u_c(c_t,l_t)} = w_t$ to be met on a balanced growth path where, as shown above, c and w grow at the same rate, and now l must be constant. It turns out that these conditions are met if and only if the utility function is of the form $u(c) = \frac{(cg(l))^{1-\sigma}-1}{1-\sigma}$, where g is strictly increasing and such that $cg(l)$ is strictly quasiconcave.[10] It is again straightforward to show the "if" part but somewhat more demanding to show the "only if" part; we do the latter in Chapter 8.

In conclusion, we have now arrived at a utility-function specification that is (the only one) consistent with choosing a constant saving rate and constant labor supply in the long run. The precise population structure and the length of households' lives can, however, satisfy a variety of assumptions and our main two applications will be the representative-agent dynasty and the simplest overlapping-generations model.

---

[9]To understand the $\sigma = 0$ case, take the limit as σ goes to 1 but use l'Hôpital's rule. One obtains $\lim_{\sigma\to 1} \frac{d(c^{1-\sigma}-1)/d\sigma}{d(1-\sigma)/d\sigma} = \lim_{\sigma\to 1} \frac{-c^{1-\sigma}\log c}{-1} = \log c$.

[10]This requires that $1/g(l)$ is strictly convex.

2.6 Preview of the whole text

The section above introduces an overall macroeconomic framework that is quantitatively oriented and that will then be put to work to analyze a number of applied macroeconomic questions. This introduction, however, motivated the overall framework rather than provided a full analysis of it, and much of the overall text is really about understanding the basic framework more fully, since so much of modern macroeconomics is built around it. To begin with, we need to study the Solow model in some more detail; this will be the topic of Chapter 3. In particular, it is important to understand its convergence properties. The Solow model can be viewed as a core model for understanding growth, even though it punts on why decisions on saving rates and working time are made the way they are made (i.e., they are approximately constant), and the insights from the large theoretical and empirical literatures that followed the seminal studies on endogenous technological change and human capital formation by Romer and Lucas in the 1980s actually have left the Solow model in rather good standing. As indicated above, the Solow model can also be used to organize much of the business-cycle research, so a more thorough analysis of the Solow model will also constitute the beginning of an analysis of short-run fluctuations.

Chapter 4 will cover optimization more fully. The discussion above merely uses microeconomic principles, such as to say that the marginal rate of substitution should equal the relative price, and even though the stated conditions will hold as part of optimal behavior, they leave open how the associated dynamic maximization problem is fully solved. Here, there are two different methods. The first one sequential is in nature, i.e., it looks for the sequences (of consumption and saving, and labor and leisure) that maximize a utility function defined over sequences, subject to some constraints. For this method, the case where the time horizon is literally infinite deserves some special treatment. The second method is dynamic programming, where the object of choice is a function (such as a consumption function or a labor-supply function), and dynamic programming has become a standard tool in macroeconomics, especially because it is often useful, if not indispensable, when the models become more complex and need to be solved numerically.

Chapter 5 looks at another very important methodological tool for all of modern macroeconomics: how market allocations—decentralized equilibria—are defined and analyzed. How to define of a competitive equilibrium may seem like a footnote but is actually a crucial element of the analysis, especially in dynamic models with some complexity: the definition specifies what the model really is, in a tight logical way, and often setting up a proper equilibrium definition is a more important element of understanding a model than actually solving for some implied equilibrium behavior. The aim here is to provide equilibrium definitions that are as compact as possible but at the same time logically and mathematically tight enough that they can be viewed as a set of instructions for a smart mathematician or programmer to follow and implement. We will look at three kinds of equilibrium definitions—two based on sequential methods and one based on dynamic programming—which will all deliver the same market allocations but emphasize different aspects of the market mechanism and therefore different applications may be best analyzed with different equilibrium concepts.

In Chapter 6, the focus is on uncertainty, a key element of business-cycle analysis but also a central element in the study of asset prices and many aspects of individual choice. Uncertainty, in some ways, is conceptually a straightforward addition to the framework, so part of the purpose is precisely to make this clear. However, solving models with uncertainty poses special challenges and another purpose is to discuss methods to this end.

The purpose of Chapter 7 is to analyze the performance of markets. I.e., here we look at whether there are welfare theorems that may apply in the simplest versions of the macroeconomic model. Almost all research in the area today is about frictions in the macroeconomic machinery, so the focus here on welfare theorems might seem misguided. However, a serious understanding of a model with frictions—as well as the analysis of what policy instruments might be helpful for dealing with the frictions—requires a solid understanding of how markets work without the frictions; often (but not always), a desirable policy is precisely to move the economy toward the frictionless case. This chapter will study two basic models: the dynastic, representative-agent model and the overlapping-generations model, which turn out to have very different efficiency properties. The presentation of the overlapping-generations economy is also important in its own right, since it is one of the key frameworks used in applied macroeconomics.

Chapter 9 begins the more applied part of the text with an analysis of economic growth. It first looks at the optimizing growth model—i.e., the Solow framework with optimal consumption-saving choices—and discusses convergence in that context. It then discusses other channels for long-run growth than capital accumulation: endogenous technological change and human capital accumulation. This chapter thus summarizes some key modeling contributions in the literature that began in the 1980s. The chapter then discusses the growth data from a variety of perspectives and ends with a summing up about where the current research stands.

In Chapter 10, business cycles are studied, first by describing some facts and then by looking at some models. These models will have the basic form described in the earlier sections but the emphasis now is how "shocks" hit the economy and how these shocks then propagate—influence, though various mechanisms, various macroeconomic outcomes—over time. Here, a key omission is money: the present text merely aims to look at real factors. This is not an implicit statement about the relevance of monetary economics but rather reflects the fact that the analysis of monetary economies will always involve an underlying real model and that the monetary elements are a form of friction—for example, money is used because transactions cannot be carried out costlessly, or monetary policy has real effects because prices cannot be changed costlessly. Because this chapter will not contain monetary analysis, it is necessarily an incomplete discussion of business cycles, and the interested reader must then consult other textbooks. However, many of the important mechanisms in models with money are studied in the chapter so, on the other hand, the material ought to be valuable for anyone interested in monetary economics.

In Chapter 11, the labor market is assumed to be working frictionlessly, and hence unemployment is abstracted from. A key purpose of Chapter 11 is to introduce the most commonly used model for understanding unemployment: the Diamond-Mortensen-Pissarides model of search and matching. However, the chapter begins by looking at the frictionless labor market in some more detail, by discussing labor demand and labor supply. This model may appear as not so closely related to the basic macroeconomic framework used in the rest of the text. However, the basic Pissarides (1985) model—which is the simplest general-equilibrium version of the search/matching model—really can be viewed as a being special case of the growth model, one with linear utility, of course with the labor-market friction added. This chapter also discusses some challenges to the basic search/matching model, such as the "Shimer critique".

Chapters 12–14 study other applications, mostly to derive some core results. Chapter 12 studies asset prices: it describes the Lucas asset-pricing model and discusses features of asset-pricing data and various asset-pricing "puzzles", such as the difficulty of the main model to explain why the equity premium is so large, why the riskfree rate is so low, and why prices are so volatile and, over the medium run, seemingly predictable. Chapter 13 looks at some key results on tax policy, such as Ricardian equivalence (or lack thereof), tax smoothing, and why in many economies it is efficient to set capital taxes equal to zero in the long run. Chapter 14, finally looks at some aspects of consumer inequality. It first looks at conditions under which some elements of inequality lead to a "reduced form" with a representative agent. It then looks briefly at sources of inequality such as wage differences across consumers and their implications for macroeconomic analysis.

[OTHER OMISSIONS, SO FAR, IN THE AVAILABLE TEXT: CREDIT-MKT FRICTIONS, INVESTMENT WITH ADJUSTMENT COSTS (TOBIN'S Q), VAR ANALYSIS, FISCAL MULTIPLIERS, BEHAVIORAL APPROACHES (E.G., THE LAIBSON MODEL)...]

Chapter 3

The Solow model

The present chapter will cover Solow's model in somewhat more detail, with emphasis on the convergence result and the applicability of the general framework. We will also, for the first time in the text, *calibrate* a model, i.e., assign parameter values and then use the model quantitatively to address questions of interest. Section 3.1 thus goes through the basic model and its convergence result and Section 3.2 goes through a number of applications.

3.1 The model

Recall that the basic model consists of a three simple equations:

$$C_t + I_t = Y_t = F\left(K_{t,} L\right) \tag{3.1}$$

$$I_t = K_{t+1} - (1-\delta) K_t \tag{3.2}$$

$$I_t = sF\left(K_t, L\right). \tag{3.3}$$

The resource constraint, equation (3.1), reminds us that government spending is abstracted for, or equivalently can be thought of as subsumed in C_t. Equation (3.2) describes capital accumulation and equation (3.3) is the only behavioral equation, stating that the investment-to-output ratio, or rate of saving, is constant over time at t. These equations together form a complete dynamic system—an equation system defining how its variables evolve over time—for some given F. That is, they tell us, in principle, what $\{K_{t+1}\}_{t=0}^{\infty}$ and $\{Y_t, C_t, I_t\}_{t=0}^{\infty}$ will be, given any initial capital value K_0.

In order to analyze the dynamics of the economy, we now make some more detailed assumptions.

- $F\left(0, L\right) = 0$.

- $F_K\left(0, L\right) > \dfrac{\delta}{s}$.

- F is strictly concave in K and strictly increasing in K.

- $\lim_{k \to \infty} sF_K(K, L) + 1 - \delta < 1.$

An example of a function satisfying these assumptions, and that will be used repeatedly in the course, is the Cobb-Douglas function $F(K, L) = AK^{\alpha}L^{1-\alpha}$ with $0 < \alpha < 1$. A is a productivity parameter usually referred to as TFP: total-factor productivity. As pointed out earlier, and α and $1 - \alpha$ will, under perfect competition in the input market, be equal to the capital and labor shares of income, respectively.

The law of motion equation for capital may be rewritten as

$$K_{t+1} = (1 - \delta) K_t + sF(K_t, L).$$

and this equation is useful to describe graphically. Thus, mapping K_t into K_{t+1} graphically, Figure 3.1 can be used to analyze the dynamics of capital in this economy.

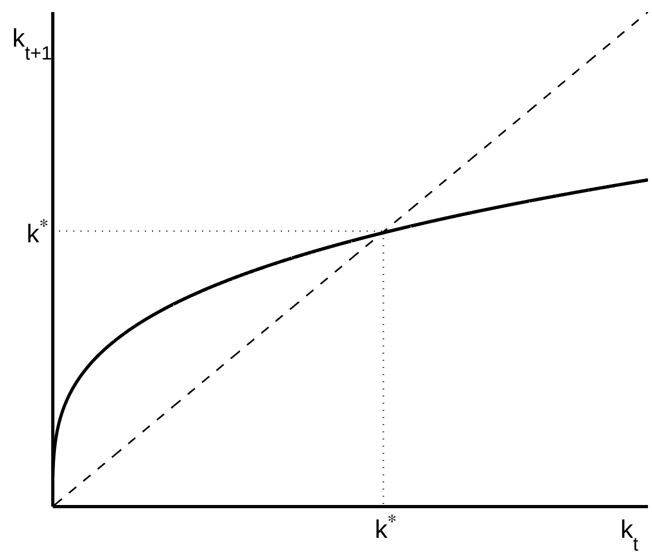

Figure 3.1: The accumulation equation and the steady state

Using the assumptions in order, the accumulation function starts at zero (the first assumption above) with a slope that exceeds 1 (the second assumption), is strictly increasing and strictly concave (follows from the third assumption), and has an asymptotic slope that is strictly below one (the last assumption). Thus, graphically, these assumptions make it clear that the accumulation function has a unique positive intersection with the 45^o line.[1] The intersection is the steady state of the dynamic system, or what might be labeled a stationary

[1] Of course, this statement can easily be proved formally but as in many similar cases throughout the text, a "graphic proof" will be deemed sufficient.

point: if the economy starts there, it remains there. The steady state, which we denote K^*, thus must satisfy

$$\begin{aligned} K^* &= (1-\delta) K^* + sF(K^*, L), \text{ or} \\ \delta K^* &= sF(K^*, L). \end{aligned}$$

Now it can be verified that the dynamic system exhibits *global and monotone convergence*, as described in Figure **??**: as long as the initial capital stock is strictly positive, the path for capital will move monotonically toward the steady state. This result is clear from the figure, since the use of the accumulation function and the 45^o line easily allow us to depict the dynamics, but it can also be proved formally as follows:

Theorem 3.1 $\forall K_0 > 0$, $K_t \to K^*$.

Proof. Beginning with the case $K_0 > K^*$, we must first show that $K^* < K_{t+1} < K_t$ $\forall t \geq 0$. Since $K_{t+1} - K_t = sF(K_t, L) - \delta K_t$, for the second inequality we must show that $sF(K_t, L) - \delta K_t > 0$. Noting that $sF(K, L) - \delta K$ is zero for $K = 0$ and for $K = K^*$, strictly concave, and increasing at 0, this must be true at least for K_0, since K_0 is positive and below steady state. To show that $K_{t+1} > K_t$ for all t, one must proceed by induction and hence show that K_{t+1} is also below steady state. Since $(1-\delta)K + sF(K, L)$ is a strictly increasing function and equals K^* as $K = K^*$, it must be strictly less than K^* for $K < K^*$. Thus, we have established that K_t satisfies the inequalities above. Similarly, we can establish that if $K_0 < K^*$ it must be that $K^* > K_{t+1} > K_t$ $\forall t > 0$. To formally show convergence, use the theorem that says that any monotonic and bounded sequence has a limit. Thus it must converge. The only possible stationary point, moreover, is the unique solution to $\delta K^* = sF(K^*, L)$. ■

Let us now briefly dwell on the relevance of the above assumptions for the convergence result, in part to better understand the result and in part because some of this discussion will come back as potentially relevant in applications. So suppose, for example, that the production function is not strictly concave. One way in which this could occur is that there is a minimum requirement on capital for any output to be generated—$F(K, L)$ would be zero for some lower range of capital levels—but that over the remaining range the production function is strictly increasing and strictly concave. Then, if there is one positive steady state, there would have to be a second one, and global convergence would not hold. Dynamics would still be monotone but for low enough initial capital stocks this monotonic convergence would be toward zero. It is of course also possible to imagine other kinds of violations of strict concavity, leading to multiple steady states.

Suppose instead that the second or fourth assumptions were violated; then there would not be any positive steady state and the economy's capital stock would either converge to zero or go toward infinity, in each case monotonically.

One can imagine much more complex dynamics, for example oscillatory or even chaotic behavior.[2] But this requires that F be decreasing over some range, and such a property

[2]Chaotic behavior can, for example, occur when the mapping from K_t to K_{t+1} is "tent-shaped".

would be difficult to motivate on the microeconomic level. On the macroeconomic level one could imagine strong negative externalities of capital—such as from pollution—but it would seem difficult to imagine that these externalities are strong enough so as to outweigh the positive marginal product of capital.

3.2 Applications

The Solow model was constructed in order to address issues of long-run growth. We will therefore briefly discuss how the model can be used for addressing growth issues. We will then turn to how it can be used to analyze short-run topics.

3.2.1 Growth

As we have seen above, the Solow model—by construction—reproduces the main long-run facts for the macroeconomic variables. However, it has predictions beyond that. One set of predictions regards the steady state, or the balanced growth path more generally, and another set of predictions concern the model's convergence properties. We will address each in turn.

Steady states

Let us first extend the model slightly to include population growth: we think of total population, and the total labor input, as growing at a constant gross rate of γ_n per year. It is then straightforward to go through the transformation of variables as before, but now by defining capital also in per-capita terms, $\hat{K}_t = K_t/(\gamma_n\gamma)^t$, to arrive at

$$\gamma_n\gamma\hat{K}_{t+1} = sF(\hat{K}_t, L) + (1-\delta)\hat{K}_t.$$

Thus, the steady state, or equivalently the balanced-growth, level of capital is now determined from

$$(\gamma_n\gamma + \delta - 1)\hat{K}_{ss} = sF(\hat{K}_{ss}, L).$$

Equivalently, because F is homogeneous of degree 1, this equation can be written $\gamma_n\gamma+\delta-1 = sF(1, L/\hat{K}_{ss})$, so comparative statics are really easy. The per-capita capital level in steady state is thus higher the lower is γ_n and δ and the higher is s and L. We can interpret these features by imagining a group of countries, all operating as closed economies, i.e., without trade, but all with labor-augmenting technology growing at the same rate γ (perhaps because knowledge about technology flows across borders without the need for trade). Then these comparative statics would translate into relative positions for capital per capita. There are straightforward measures of population growth and the rate of saving from national accounts. Let us interpret L as the quality of human capital in the country, perhaps as given by education. The parameter δ, finally, is harder to obtain direct measurements for; it can be interpreted either as physical depreciation or as some form of economic obsolescence parameter, but with the latter interpretation it would be hard to entirely disentangle it from technical progress more generally. So let us now look among the countries to see what their relative positions would be. The comparative-statics result say that, in terms of observables, countries with higher population growth would have lower capital per capita—intuitively since any given saving amount of saving per capita at t translates into less capital per person the higher is population growth between t and $t+1$—as would countries with lower saving rates or lower education per capita. If we use the fact that $\hat{Y}_{ss} = F(\hat{K}_{ss}, L)$, it is easy to see that the same will hold for output per capita. Do these predictions hold in the data?

Qualitatively, we see in the following pictures that these predictions are borne out, at least in terms of unconditional correlations. Of course, these correlations cannot be directly

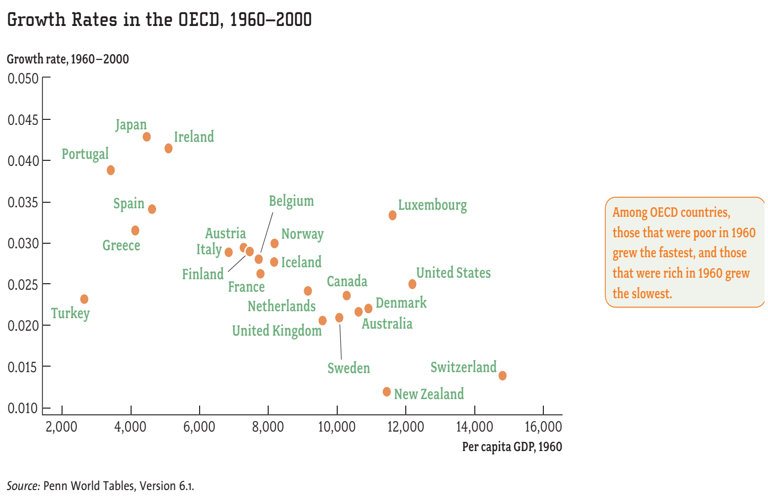

Figure 3.2: a random picture

interpreted as in terms of the causality suggested by the Solow model; perhaps higher income per capita instead changes the attitude toward children—say, via the quantity vs. quality theory of fertility. This difficulty in establishing causality of (especially long-run) relationships in macroeconomic data is unfortunately one that we cannot easily overcome. Nevertheless, we will try to draw out predictions and at least look at simple correlations in the data, if nothing else for the purpose of insisting that the models, however interesting to study, are not an end in themselves: they are meant to help us interpret the real world.

Taking the model more seriously—*calibration* One can also take the model one level more seriously and ask to what quantitative extent it can account for the correlation patterns. To do this, however, one often must specify the model more fully. In the present case, for a fully specified Solow model, one would need specific parameter values for γ_n, γ, δ, and s, as well as a functional form (with its parameters) for F. As we argued above in section 2.4, the Cobb-Douglas function is used in most applications because it allows us to match the (near-)constancy of the capital income share in the data. So let us use available data to "calibrate" the model more precisely to the growth facts in Chapter 2 and then use it to generate comparative statics that can be compared quantitatively to the data. We thus need to decide on the length of a time period; let it be one year. The depreciation rate is hard to measure directly; in business accounting equipment and structures are written off at rates that differ by type of specific capital and often the write-offs are linear, i.e., not geometric as in the model. Given at least these accounting practices, a yearly average depreciation rate of around 0.1 is probably reasonable. [NEED DATA SOURCES HERE.] Population growth and saving rates differ across countries; most developed countries are in the vicinity of $\gamma_n = 1.005$ and $s = 0.3$. Long-run growth has been around, or somewhat below, 2%, so let us set $\gamma = 1.02$. Finally, the level of the capital share varies depending on how self-employment income is categorized but suppose we take U.S. data and regard the capital share out of self-employment income as the same as for the rest of the economy. Then we arrive at a share of around 30%. So with $F(k, l) = Ak^\alpha l^{1-\alpha}$ we then set $\alpha = 0.3$. It remains to choose A and L. The choice of A is immaterial; its value will influence the level of capital in steady state but it will not influence any observable ratios such as the capital-output ratio: the capital-output ratio is directly given by $s/(\gamma_n\gamma + \delta - 1)$, independently of A (and α). Noticing this, we also see that an alternative way of calibrating δ is to use values for K constructed in the national statistics and relate them to GDP (as an average over some period of time), thus making it possible to back out a value for δ. For a value of $\delta = 0.1$ we obtain $K/Y \approx 2.4$, which is in line with data.[3] So let us then select $A = 1$. L, similarly, can also be normalized (it enters multiplicatively with A), so let the level of L in the U.S. be set to 1.

Armed with numerical parameter values and the key steady-state equation

$$\hat{Y}_{ss} = \left(\frac{s}{\gamma_n\gamma + \delta - 1}\right)^{\frac{\alpha}{1-\alpha}} L,$$

which is based on substituting capital for $\hat{K}_{ss}/\hat{Y}ss = s/(\gamma_n\gamma + \delta - 1)$ in the production function, we are now ready to take the model more seriously. Thus, we can for example compute the variations in relative output that would result from population growth varying between $\gamma_n = 1$ and $\gamma = 1.02$ and draw this line through the data points in the previous graph [GRAPH HERE]. We can also vary L and see how output is affected. As we will argue in more detail in the growth chapter below, one year of additional education is widely considered to contribute to labor productivity by about 10%. Thus we can generate variation in output by looking at education levels that go between, say, average the U.S. level and 5 years less (among OECD countries, the variation is much smaller). Thus, we can similarly add a line predicted by the theory in the education-output graph [GRAPH HERE]. Lastly, we can vary saving rates between, say 0.2 and 0.4, to generate long-run output differences, as plotted in the figure on investment-output ratios and GDP [GRAPH HERE].

Convergence, or lack thereof

The Solow model also has more detailed predictions for convergence; in particular, one can use it analyze the convergence *speed*. As an illustration, consider the simple Cobb-Douglas

[3] Note here that had we chosen, say, a time period to be a quarter, GDP would be 4 times smaller and K/Y correspondingly 4 times larger; of course, γ_n, γ, and δ would then have to be adjusted down as well.

case. In that case, α—the capital share—determines the shape of the law of motion function for capital accumulation. If α is close to one the law of motion is close to being linear in capital; if it is close to zero (but not exactly zero), the law of motion is quite nonlinear in capital. In terms of Figure 3.1, an α close to zero will make the steady state lower, and the convergence to the steady state will be quite rapid: from a given initial capital stock, few periods are necessary to get close to the steady state. If, on the other hand, α is close to one, the steady state is far to the right in the figure, and convergence will be slow.

Locally, around the steady state point, Figure **??** illustrates that the slope of the accumulation equation is directly informative of the convergence speed: a slope closer to one indicates slower convergence. Formally, we can therefore define this speed, defined through a paramter λ, by defining a gap, $K_t - K^*$, between the current capital stock and the steady-state level and then using

$$K_{t+1} - K^* = (1 - \lambda)(K_t - K^*);$$

λ measures what fraction of the initial gap is closed in one time period, a measure that does not depend on time for a linear function. Thus, $\lambda = 0$ is the case where the gap is not changing at all, so the speed of convergence is zero, and $\lambda = 1$ is the opposite extreme case, where the speed of convergence is infinite since the gap is closed fully in one period.

For the Solow model, what parameters determine λ? We can see that the equation $K_{t+1} - K^* = (1 - \lambda)(K_t - K^*)$ can be represented rewritten and reinterpreted as a first-order Taylor expansion of the capital accumulation equation. Thus, $1 - \lambda$ corresponds to the derivative of $sF(K_t, L) + (1 - \delta)K_t$ with respect to K_t, evaluated at the steady state. Using the Cobb-Douglas case, we obtain $1 - \lambda = \alpha s A(K^*/L)^{\alpha-1} + 1 - \delta$, and given that $sA(K^*/L)^{\alpha-1} = \delta$ in steady state (looking at the case without either technology or population growth), we obtain $1 - \lambda = \alpha\delta + 1 - \delta$, so $\lambda = \delta(1 - \alpha)$.[4] Thus, we have a higher convergence speed the higher is δ and the lower is α.

So does convergence hold in the data? It is important to recognize here that convergence to a unique stationary point occurs only for each unique set of parameter values; if, for example, s is higher in one country than in another, their levels of capital or output levels would not be expected to converge. Therefore, one can take a rough look at whether convergence holds or not by restricting attention to "similar" countries. So consider the OECD, and consider looking at whether, within this group of countries those with relatively high initial capital stocks grow faster subsequently. Figure 3.2.1 illustrates convergence, thus qualitatively at least confirming the predictions of Solow's growth model. We will return in the chapter on growth to the issue of convergence and show how convergence can be examined more formally.

With the calibration carried out in the previous section, we can however also assess how well the Solow model matches the data quantitatively, from the convergence perspective. The parameter vector $(\gamma_n, \gamma, \delta, s, \alpha) = (1.005, 1.02, 0.10, 0.3, 0.3)$ implies that λ equals around 0.014. In the data, we can obtain λ by looking at $\frac{K_{t+1}-K_t}{K^*}$ as a function of K_t/K^*.

[4] With population and technology growth, we obtain $1 - \lambda = \alpha(\gamma_n\gamma + \delta - 1) + 1 - \delta$, so $\lambda = \alpha(\gamma_n\gamma - 1) + \delta(1 - \alpha)$.

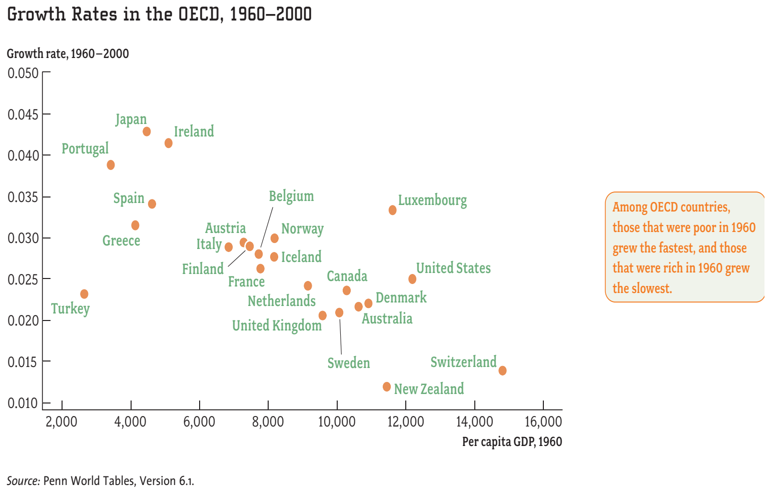

Figure 3.3: Convergence within the OECD

One special case of the Solow model deserves separate interest: the case where the production function is linear in capital—when α equals one. Then there is generically no positive steady state.[5] Suppose that $sA + 1 - \delta$ exceeds one. Then over time output would keep growing, and it would grow at precisely rate $sA+1-\delta$. Output and consumption would grow at that rate too. The technology in this case is referred to as the "Ak" production function and is central for understanding the mechanics of the concept "endogenous growth", which will be discussed in much more detail in Chapter 9 below. Endogenous means something like "nontrivially determined", at least in the minimal sense that different types of behavior correspond to different growth rates. Clearly, in this case, the saving rate influences the growth rate directly, even to the point where a rate that is very low will even make the economy shrink—the case where $sA + 1 - \delta$ is below one. Keeping in mind that savings rates are not just set by people but arguably influenced by government policy, such as taxation, this means that there would be a *choice*, both by individuals and government, of whether or not to grow, and more generally at which rate to grow. Thus, small policy differences across countries can imply very large (and growing) long-run consequences for relative performances.

The "Ak" model of growth emphasizes physical capital accumulation as the driving force of prosperity. It is not the only way to think about growth, of course. For example, one could model technological change more carefully and be specific about how productivity is enhanced over time via explicit decisions to accumulate R&D capital or human capital—learning. We will return to these different alternatives later, in the chapter on growth.

[5]When $sA+1-\delta$ happens to equal 1, there is a continuum of steady states, as the accumulation equation coincides with the 45^o line.

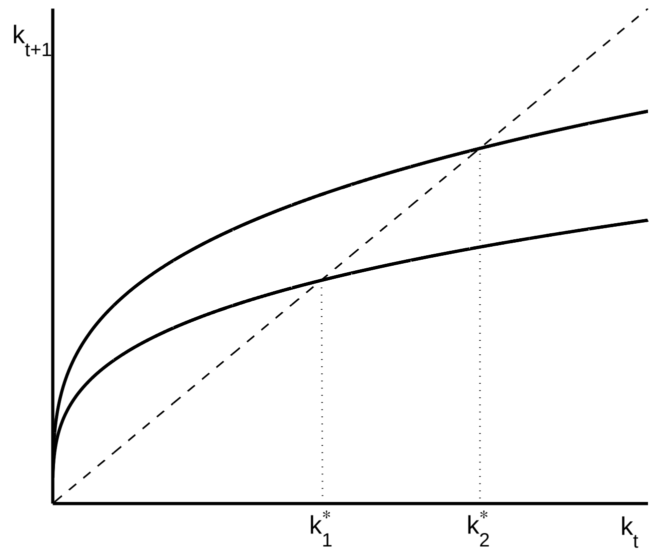

Figure 3.4: Random productivity in the Solow model

3.2.2 Business cycles

Many modern studies of business cycles—real as well as monetary—also rely on a close relative of the Solow model. How can Solow's framework be used as a business-cycle setup?

Technology and other shocks

As a first example, assume that the production technology will have a stochastic component affecting the productivity of factors. This would then be a very simple version of what has been labeled the real business cycle model. So assume, for example, that production is $A_t \hat{F}(K_t, L)$ where A_t is iid (identically and independently distributed over time) and can take on two values only: A_H and $A_L < A_H$. Retaining the assumption that savings rates are constant, we have what is depicted in Figure 3.4.

It is clear from studying this graph that as productivity realizations turn between high and low, output and total savings fluctuate. Will there be convergence to a steady state? In the sense of constancy of capital and other variables, steady states will clearly not be feasible here. However, another aspect of the convergence in deterministic model is inherited here: over time, initial conditions (the initial capital stock) lose influence and, eventually, after sufficiently long, the stochastic process for the endogenous variables will settle down and become stationary. Stationarity here is a statistical term, one that we will discuss in some more detail in Chapter 6 below. Statistical stationarity is an important element of the analysis of past data, and it is the natural extension of the focus on balanced growth in the case of certainty. The stochastic Solow model delivers a joint process for exogenous productivity process and capital, which is endogenous, and finding (and even defining) the associated stationary distribution is rather involved. Defining and finding a stationary distribution for an exogenous variable, such as productivity here, is relatively straightforward, and many applications in the text (such as those on asset pricing in Chapter 12 below), will not go beyond this point. One element of stationarity in the case of the Solow model is that there will be a smallest compact set of capital stocks such that, once the capital stock is in this set, it never leaves the set: the "ergodic set". In the figure, this set is determined by the two intersections with the 45^o line, K_1^* and K_2^*. In the case with an iid productivity process, the stationary distribution for capital will, moreover, be independent of productivity, and it will have a shape like that depicted in the figure. This shape is, in general, hard to compute directly; in practice, one would simulate the model on the computer and draw a histogram based on the simulations.

Second, will a model of this sort be able to account for the fluctuations in aggregate variables? The setup described here generates data for output, consumption, investment, and capital. Looking at the three first variables, we can look at (i) comovements and (ii) relative variabilities. It is clear, since $I/Y = s$ and $C/Y = 1 - s$ in the Solow model, that (i) there is one-for-one comovement of all three variables and (ii) that their percentage fluctuations are identical. The first of these observations is very much in line with overall data: the different expenditure categories comove rather strongly. The second, however, is strongly at odds with data: investment is much more, and consumption much less, volatile than GDP in percentage terms. So the Solow model does not allow us to account for relative volatilities. However, there are reasons to think that as productivity fluctuates, the optimal, or market-induced, saving rate ought to fluctuate. In the sections on optimization, as well as in the chapter on business cycles, we will indeed derive optimal saving behavior in the presence of productivity uncertainty. Similarly, labor supply ought perhaps not be constant if subject to choice. Thus, in the optimizing models in the remainder of the text, we will in general have both the saving rate and labor input move with the shocks, as well as with the capital stock for any given shock value. Thus, the optimizing model will be of the form $K_{t+1} = s(K_t, A_t)A_tF(K_t, L(K_t, A_t)) + (1 - \delta)K_t$, where the saving rate and labor are now functions of the current capital stock as well as of the productivity level.[6] In terms of our graphical analysis, thus, the shapes of the two functions are more non-trivially determined, but the actual dynamics can still be followed by following the two functions and the shock realizations. One basic finding of the optimal growth model under uncertainty below will actually be that the relative fluctuations of consumption, investment, and output predicted by the model are quite similar to those found in the data.

Business-cycle models can of course involve other shocks. In a straightforward extension of the basic model, consumption and investment are still perfect substitutes but the coefficient of substitution is random: $C_t + \nu_t I_t = Y_t$. A market interpretation of this model will allow us to interpret ν_t as the relative price of investment, which when allowed to be

[6]There are, actually, very particular assumptions on preferences and technology for which both the saving rate and labor supply remain constant despite recurring shocks and for all values of capital; however, these assumptions are too strong to be taken seriously in the business-cycle context.

stochastic can contribute to the fluctuations in the macroeconomic variables. The implied Solow model would thus deliver a capital accumulation equation, assuming that $\nu_t I_t = sY_t$, $K_{t+1} = \frac{s}{\nu_t}F(K_t, L) + (1-\delta)$, that is formally identical to that with productivity changes described above, though output here will still be non-random and equal to $F(K_t, L)$. Variations of this model are also common in the literature on how technological change can cause fluctuations in output. Another type of shock is a shock to government spending.

Supply vs. demand shocks

Comovements in macroeconomic variables in the technology-based story behind cycles comes from the notion that the overall resource availability goes up and down with productivity; as long as at least some of the added new resources are used on both consumption and investment, there will clearly be comovement. In most undergraduate textbooks, however, where keynesian analysis is dominant, other shocks are in the foreground: demand shocks and shocks to monetary policy. It is, however, possible—as indicated in the introductory chapter—to also think about demand shocks in the context of the Solow model. Thus, suppose there is less than full utilization of inputs; to make the illustration simple, suppose there is less than full employment, so that labor input equals $L(1-u_t)$, where unemployment, $u_t \in [0, 1]$, has a time subscript, since we can imagine that it fluctuates over time. Imagine further that, like in the Solow model, $I_t/C_t = s/(1-s)$ at all times. Then from the equations so far, it is feasible to have $C_t \in [0, (1-s)F(K_t, L)]$, i.e., consumption can be anything between zero—a collapsed economy with 100% unemployment—and $1-s$ times full-employment output. Now turn this around to say that consumption is actually exogenous: any number in the given range is possible, and this value of consumption will imply a level of unemployment. Thus, output is demand-determined.[7] For this story to make sense on a microeconomic level, we would need to describe market frictions that make under-utilization of inputs possible; why don't wages fall until there is full employment, for example? And what model of consumption behavior allows "whims" of this sort to be of macroeconomic relevance? These are million-dollar questions in macroeconomic research, and a quick summary of the research to date is that (i) the model framework presented in Chapter 11, the Diamond-Mortensen-Pissarides model of search and matching in the labor market, offers a coherent description of markets that imply unemployment, but (ii) whereas there is significant research ongoing on explaining consumption, there is as of yet no generally accepted framework that provides a consistent theory of consumption/demand shocks, linked to unemployment or more generally under-utilization of resources. Some theories in this direction will, however, be discussed toward the end of Chapter 10. For now, however, we can envision a version of the Solow model with an accumulation equation given by $K_{t+1} = (1-\delta)K_t + sF(K_t, L(1-u_t))$, with u_t moving exogenously over time.[8]

As for monetary policy shocks, and more generally a framework where monetary non-

[7]On this level of generality, this formulation does not express the kinds of multiplier effects that are a central elements behind the IS curve of the keynesian analysis. They would be straightforward to introduce, however.

[8]Krusell and Smith (1998) uses this kind of setting.

neutralities play an important role for macroeconomic variables, there is by now a large literature. However, that literature can also be described, in a reduced-form/short-hand fashion, with the kind of utilization variation just discussed. For readers interested in these models, a very useful first reference is Galí's textbook.

Impulse responses

A very common tool for illustrating how shocks affect the economy—be it technology shocks or other shocks—is the *impulse response diagram*. The typical background scenario is that the economy is at steady state, and that an exogenous change occurs. Suppose, as an illustration, that there is an exogenous increase in total-factor productivity, A_t, given the production function $A_t K_t^{\alpha} L^{1-\alpha}$, with an otherwise standard Solow model without recurring technological or population change. Suppose, moreover, that the change in productivity has the following pattern over time: it is 1% higher than the initial steady-state value in the initial period, ρ% higher the next period (with $\rho \in (0,1)$), ρ^2% higher the period after that, and so on. That is, the productivity increase is largest initially and then peters out slowly and converges to its original, steady-state value. It is straightforward to compute the dynamics for capital and output by just iterating on the accumulation equation $K_{t+1} = sA_t K_t^{\alpha} L^{1-\alpha} + (1-\delta)K_t$.[9] The resulting paths, based on the basic calibration of Solow's model from before, for capital and output are depicted in Figure **??** below, expressed as percentage changes in capital and output, relative to steady state.

[IMPULSE-RESPONSE FIGURE HERE]

In the Solow model, behavior is mechanical and represented by s and L being constant, but in an optimizing growth model these variables would in general respond to the productivity increase. We will discuss how impulse responses in such models are generated in Chapter 10 below. In an optimizing model, it would then also be important to make clear that the productivity change is unexpected: for if it were expected, the assumption of being in a steady state initially would be hard to justify. This also suggests that one should study shocks that are recurring, i.e., shocks that, one by one, might be impossible to predict but that agents would have an overall awareness of; they might, for example, take precautions in expecting shocks to occur in case they are averse to risk. Thus, in order to fully understand how shocks influence the macroeconomy one needs to complement impulse-response analysis with simulations where shocks are recurrent and the macroeconomic decision makers take this into account in their decisions.

Linearization and impulse responses In actual applications, often the impulse responses are approximated by a linearization technique. I.e., a first-order Taylor expansion of the capital accumulation equation is derived with respect to capital and productivity, and then the linear system is used to generate the capital path; similarly, output as a function of

[9] Normalizing the steady-state value of A to 1, we would have $K_0 = \left(\frac{s}{\delta}\right)^{\frac{1}{1-\alpha}} L$ and then $K_1 = 1.01sK_0^{\alpha}L^{1-\alpha} + (1-\delta)K_0$, $K_2 = (1+0.01\rho)sK_1^{\alpha}L^{1-\alpha} + (1-\delta)K_1$, $K_3 = (1+0.01\rho^2)sK_2^{\alpha}L^{1-\alpha} + (1-\delta)K_2$, and so on.

productivity and capital is also linearized. Linearization is practical since it allows closed-form expressions that can be easily programmed. The linearized responses are also plotted in Figure **??**, again in percentage terms. It is clear from the resulting impulse responses that a linear approximation produces a highly accurate representation of the non-linear dynamics for the calibration used. One always has to be aware, of course, that larger shocks may produce larger approximation errors—for infinitesimal shocks the linear approximation is exact—but for most calibrated models analyzed in this text, the first-order approximation turns out to be rather accurate.

To see how linearization works in concrete terms, consider the path for capital first. Capital follows $K_{t+1} = sA_tK_t^{\alpha}L^{1-\alpha} + (1-\delta)K_t$, where $A_t = \rho^t A_0$ and A_0 is shocked away from its steady-state value, 1 (e.g., $A_0 = 1.01$). Also, let us aim at producing the impulse responses in percentage terms—since approximations are often carried out this way—and hence we define $\tilde{K}_t = \log K_t$ so that $d\tilde{K}_t = dK_t/K_t$, i.e., $d\tilde{K}_t$ expresses a percentage change in capital. Similarly, $d\tilde{A}_t = dA_t/A_t$. Taking total differentials and using our definitions, we thus obtain

$$K_{t+1}d\tilde{K}_{t+1} = sK_t^{\alpha}L^{1-\alpha}A_t d\tilde{A}_t + \alpha s A_t K_t^{\alpha-1}L^{1-\alpha}K_t d\tilde{K}_t + (1-\delta)K_t d\tilde{K}_t.$$

Simplifying somewhat, evaluating at steady state, where $sY = \delta K$, and dividing by K on both sides, we obtain

$$d\tilde{K}_{t+1} = \delta\tilde{A}_t + (1-\delta(1-\alpha))d\tilde{K}_t.$$

This is the linear form sought. If we begin by setting $d\tilde{K}_0 = 0$ and then solving for $\tilde{K}_1$, we obtain $\delta d\tilde{A}_0$. We can then iterate forward. Using the notation $\lambda \equiv 1-\delta(1-\alpha) < 1$ to shorten the expressions, we see that $d\tilde{K}_2 = (\rho+\lambda)\delta d\tilde{A}_0$, $d\tilde{K}_3 = (\rho^2+\rho\lambda+\lambda^2)\delta d\tilde{A}_0$, and so on; the general expression becomes

$$d\tilde{K}_{t+1} = (\rho^t + \rho^{t-1}\lambda + \cdots + \rho\lambda^{t-1} + \lambda^t)\tilde{A}_0.$$

Thus, if $\rho = 0$, so that the shock only lived for one period, the response of capital would be to decline geometrically. With a shock that is persistent, we see that the dynamics are more complicated. As time passes, we obtain more and more terms, but each term must go to zero, since ρ and λ both are less than 1. It is easy to compute the sum: factorize the largest of ρ and λ—suppose for sake of illustration it is ρ—to obtain $\rho^t\left(1+\frac{\lambda}{\rho}+\cdots+\left(\frac{\lambda}{\rho}\right)^{t-1}+\left(\frac{\lambda}{\rho}\right)^{t}\right)$. The expression in parenthesis can be written $\frac{1-\left(\frac{\lambda}{\rho}\right)^{t+1}}{1-\frac{\lambda}{\rho}}$. Thus, we finally have, for $t \geq 0$,

$$d\tilde{K}_{t+1} = \rho^t\frac{1-\left(\frac{\lambda}{\rho}\right)^{t+1}}{1-\frac{\lambda}{\rho}}\delta d\tilde{A}_0.$$

This expression goes to zero for any $\rho < 1$, since $\lambda/\rho < 1$, but more slowly than a geometric series.[10]

[10]The expression goes to zero more slowly than either ρ^t or λ^t.

Turning to output, we obtain by straightforward differentiation again that $d\tilde{Y}_t = d\tilde{A}_t + \alpha d\tilde{K}_t$, and using our result for capital we then have

$$d\tilde{Y}_t = \rho^t d\tilde{A}_0 + \alpha\rho^{t-1}\frac{1-\left(\frac{\lambda}{\rho}\right)^t}{1-\frac{\lambda}{\rho}}\delta d\tilde{A}_0 = \rho^t\frac{\alpha\delta}{\rho}\frac{1-\left(\frac{\lambda}{\rho}\right)^t}{1-\frac{\lambda}{\rho}}d\tilde{A}_0$$

for all $t > 0$ (for $t = 0$ we have $d\tilde{Y}_0 = d\tilde{A}_0$). Thus, we have solved, in closed form, for the impulse responses of capital and output in terms of the primitives of the growth model. Note that the only three parameters that matter are ρ, δ, and α; if we had linearized in terms of levels, s and l would have appeared as well.

Impulse responses and the data Impulse responses are useful for understanding the dynamics of theoretical models. For example, if one looks at the expression for $d\tilde{Y}_t$ one sees that a high α makes output move more slowly over time: it slows down the convergence back to steady state in response to shocks. This is not surprising in light of our graphical discussions of the Solow model. However, impulse responses are also very valuable for comparisons between model and data. However, how does one identify an impulse response in the data? One can go about this in a variety of ways, and these different methods will at least briefly be discussed below. A central challenge is to identify a "surprise event" like the one we just analyzed, and then to ascertain that movements subsequent to the shock are due to the shock and not due to other factors; these are highly challenging tasks since the macroeconomic data consists of variables that are interdependent, often with two-way feedback mechanisms, and where expectations—reflecting events not yet observed—can influence current outcomes, and so on. Thus, one direct approach is to try to find direct ways of convincing oneself that a shock is really both exogenous and a surprise. This is something that in macroeconomics at best can be accomplished for a subset of particular variables at particular times. One method of this sort is the "narrative approach" by Romer and Romer and used, so far, for some forms of monetary and fiscal policy shocks; in these applications, their method, in brief, was to use detailed inside knowledge about the policymaking process and to simply determine that some policy changes satisfied the desired criteria of exogeneity and unexpectedness.

Another method is to use structural econometrics: assume that the model is true and estimate its parameters by making the model fit the data as closely as possible, which could be accomplished with classical or Bayesian methods. Then analyze the impulse-response implications given the theory based on the estimated parameters. One can of course imagine structural estimation with instrumental variables, if it is possible to find an instrument that satisfies the desired criteria, but few variables have been deemed to be credible as instruments, at least for the standard macroeconomic variables.

Finally, vector autoregressions—VARs—represent an alternative approach that uses some theory but not a fully specified model. VARs, as developed by Sims (19xyz), are linear relations between a subset of macroeconomic variables and a their history, and they are very common in applied macroeconomic analysis. Section xyz below discusses them briefly, though of course much of the detailed treatment of VARs, as well as of other econometric techniques relevant in macroeconomics, are left outside the scope of the present text.

3.2.3 Using the Solow model for other applied macroeconomic questions

The studies of growth and business cycles are the most central ones in macroeconomics, and the previous discussion aims to show how the Solow model—and extensions of it that allow us to give market interpretations of this setting—is useful for conducting such studies. The setting is however also used for other applied questions. One is a perennial public-finance question: how should the government finance its spending? I.e., to what extent should it tax labor earnings versus capital income, or use consumption taxes? And when should it levy its taxes: at the same rate as the spending occurs, or more smoothly over time, or in some other manner? The answer has implications for the evolution of government debt. These questions, which are macroeconomic in nature, and should be subjected to the general quantitative discipline, can be straightforwardly addressed with the basic setting in this text; Chapter 13 looks at them. There are also questions on interest-rate determination and asset pricing that demand a setup of the kind analyzed here, but most of the asset-pricing analysis in Chapter 12 will be conducted in models with exogenous output and without capital accumulation.

Let us also mention the need to distinguish closed-economy from open-economy settings. The Solow model is set up for a closed economy and open-economy environments really call for explicit choices, e.g., regarding international borrowing and lending. Open-economy models are used, and will be treated as important special cases, in both the growth and the business-cycles chapters below.

Chapter 4

Dynamic optimization

There are two common approaches to modelling real-life individuals: (i) they live a finite number of periods and (ii) they live forever. The latter is the most common approach, but the former requires less mathematical sophistication in the decision problem. We will start with finite-life models and then consider infinite horizons.

We will also study two alternative ways of solving dynamic optimization problems: using sequential methods and using recursive methods. Sequential methods involve maximizing over sequences. Recursive methods - also labelled dynamic programming methods - involve functional equations. We begin with sequential methods and then move to recursive methods.

4.1 Sequential methods

4.1.1 A finite horizon

Consider a consumer having to decide on a consumption stream for T periods. Consumer's preference ordering of the consumption streams can be represented with the utility function

$$U\left(c_0, c_1, ..., c_T\right).$$

A standard assumption is that this function exhibits "additive separability", with stationary discounting weights:

$$U\left(c_0, c_1, ..., c_T\right) = \sum_{t=0}^{T} \beta^t u\left(c_t\right).$$

Notice that the per-period (or instantaneous) utility index $u\left(\cdot\right)$ does not depend on time. Nevertheless, if instead we had $u_t\left(\cdot\right)$ the utility function $U\left(c_0, c_1, ..., c_T\right)$ would still be additively separable.

The powers of β are the discounting weights. They are called stationary because the ratio between the weights of any two different dates $t = i$ and $t = j > i$ only depends on the number of periods elapsed between i and j, and not on the values of i or j.

The standard assumption is $0 < \beta < 1$, which corresponds to the observations that human beings seem to deem consumption at an early time more valuable than consumption further off in the future.

We now state the dynamic optimization problem associated with the neoclassical growth model in finite time.

$$
\begin{array}{rl}
\max\limits_{\{c_t, k_{t+1}\}_{t=0}^{T}} & \sum\limits_{t=0}^{T} \beta^t u(c_t) \\
\text{s.t.} & c_t + k_{t+1} \leq f(k_t) \equiv F(k_t, N) + (1 - \delta) k_t, \forall t = 0, ..., T \\
& c_t \geq 0, \forall t = 0, ..., T \\
& k_{t+1} \geq 0, \forall t = 0, ..., T \\
& k_0 > 0 \text{ given.}
\end{array}
$$

This is a consumption-savings decision problem. It is, in this case, a "planning problem": there is no market where the individual might obtain an interest income from his savings, but rather savings yield production following the transformation rule $f(k_t)$.

The assumptions we will make on the production technology are the same as before. With respect to u, we will assume that it is strictly increasing. What's the implication of this? Notice that our resource constraint $c_t + k_{t+1} \leq f(k_t)$ allows for throwing goods away, since strict inequality is allowed. But the assumption that u is strictly increasing will imply that goods will not actually be thrown away, because they are valuable. We know in advance that the resource constraint will need to bind at our solution to this problem.

The solution method we will employ is straight out of standard optimization theory for finite-dimensional problems. In particular, we will make ample use of the Kuhn-Tucker theorem. The Kuhn-Tucker conditions:

(i) are necessary for an optimum, provided a constraint qualification is met (we do not worry about it here);

(ii) are sufficient *if* the objective function is concave in the choice vector and the constraint set is convex.

We now characterize the solution further. It is useful to assume the following: $\lim\limits_{c \to 0} u'(c) = \infty$. This implies that $c_t = 0$ at any t cannot be optimal, so we can ignore the non-negativity constraint on consumption: we know in advance that it will not bind in our solution to this problem.

We write down the Lagrangian function:

$$L = \sum_{t=0}^{T} \beta^t \left[u(c_t) - \lambda_t \left[c_t + k_{t+1} - f(k_t) \right] + \mu_t k_{t+1} \right],$$

where we introduced the Lagrange/Kuhn-Tucker multipliers $\beta^t \lambda_t$ and $\beta^t \mu_t$ for our constraints. This is formulation A of our problem.

The next step involves taking derivatives with respect to the decision variables c_t and k_{t+1} and stating the complete Kuhn-Tucker conditions. Before proceeding, however, let us take a look at an alternative formulation (formulation B) for this problem:

$$L = \sum_{t=0}^{T} \beta^t \left[u\left[f\left(k_t\right) - k_{t+1} \right] + \mu_t k_{t+1} \right].$$

Notice that we have made use of our knowledge of the fact that the resource constraint will be binding in our solution to get rid of the multiplier $\beta^t \lambda_t$. The two formulations are equivalent under the stated assumption on u. However, eliminating the multiplier $\beta^t \lambda_t$ might simplify the algebra. The multiplier may sometimes prove an efficient way of condensing information at the time of actually working out the solution.

We now solve the problem using formulation A. The first-order conditions are:

$$\begin{aligned} \frac{\partial L}{\partial c_t} &: \ \beta^t \left[u'\left(c_t\right) - \lambda_t \right] = 0, \ t = 0, ..., T \\ \frac{\partial L}{\partial k_{t+1}} &: \ -\beta^t \lambda_t + \beta^t \mu_t + \beta^{t+1} \lambda_{t+1} f'\left(k_{t+1}\right) = 0, \ t = 0, ..., T-1. \end{aligned}$$

For period T,

$$\frac{\partial L}{\partial k_{T+1}} : -\beta^T \lambda_T + \beta^T \mu_T = 0.$$

The first-order condition under formulation B are:

$$\begin{aligned} \frac{\partial L}{\partial k_{t+1}} &: \ -\beta^t u'\left(c_t\right) + \beta^t \mu_t + \beta^{t+1} u'\left(c_{t+1}\right) f'\left(k_{t+1}\right) = 0, \ t = 0, ..., T-1 \\ \frac{\partial L}{\partial k_{T+1}} &: \ -\beta^T u'\left(c_T\right) + \beta^T \mu_T = 0. \end{aligned}$$

Finally, the Kuhn-Tucker conditions also include

$$\mu_t k_{t+1} = 0, \ t = 0, ..., T$$

$$\lambda_t \geq 0, \ t = 0, ..., T$$

$$k_{t+1} \geq 0, \ t = 0, ..., T$$

$$\mu_t \geq 0, \ t = 0, ..., T.$$

These conditions (the first of which is usually referred to as the complementary slackness condition) are the same for formulations A and B. To see this, we use $u'\left(c_t\right)$ to replace λ_t in the derivative $\frac{\partial L}{\partial k_{t+1}}$ in formulation A.

Now noting that $u'(c) > 0 \ \forall c$, we conclude that $\mu_T > 0$ in particular. This comes from the derivative of the Lagrangian with respect to k_{T+1}:

$$-\beta^T u'\left(c_T\right) + \beta^T \mu_T = 0.$$

But then this implies that $k_{T+1}=0$: the consumer leaves no capital for after the last period, since he receives no utility from that capital and would rather use it for consumption during his lifetime. Of course, this is a trivial result, but its derivation is useful and will have an infinite-horizon counterpart that is less trivial.

The summary statement of the first-order conditions is then the "Euler equation":

$$u'\left[f\left(k_t\right)-k_{t+1}\right]=\beta u'\left[f\left(k_{t+1}\right)-k_{t+2}\right]f'\left(k_{t+1}\right),\ t=0,...,T-1$$
$$k_0 \text{ given, } k_{T+1}=0,$$

where the capital sequence is what we need to solve for. The Euler equation is sometimes referred to as a "variational" condition (as part of "calculus of variation"): given to boundary conditions k_t and k_{t+2}, it represents the idea of varying the intermediate value k_{t+1} so as to achieve the best outcome. Combining these variational conditions, we notice that there are a total of $T+2$ equations and $T+2$ unknowns - the unknowns are a sequence of capital stocks with an initial and a terminal condition. This is called a *difference equation* in the capital sequence. It is a *second-order* difference equation because there are two lags of capital in the equation. Since the number of unknowns is equal to the number of equations, the difference equation system will typically have a solution, and under appropriate assumptions on primitives, there will be only one such solution. We will now briefly look at the conditions under which there is only one solution to the first-order conditions or, alternatively, under which the first-order conditions are sufficient.

What we need to assume is that u is concave. Then, using formulation A, we know that $U=\sum\limits_{t=0}^{T}u\left(c_t\right)$ is concave in the vector $\{c_t\}$, since the sum of concave functions is concave. Moreover, the constraint set is convex in $\{c_t,k_{t+1}\}$, provided that we assume concavity of f (this can easily be checked using the definitions of a convex set and a concave function). So, concavity of the functions u and f makes the overall objective concave and the choice set convex, and thus the first-order conditions are sufficient. Alternatively, using formulation B, since $u(f(k_t)-k_{t+1})$ is concave in (k_t,k_{t+1}), which follows from the fact that u is concave and increasing and that f is concave, the objective is concave in $\{k_{t+1}\}$. The constraint set in formulation B is clearly convex, since all it requires is $k_{t+1}\geq 0$ for all t.

Finally, a unique solution (to the problem as such as well as to the first-order conditions) is obtained if the objective is strictly concave, which we have if u is strictly concave.

To interpret the key equation for optimization, the Euler equation, it is useful to break it down in three components:

$$\underbrace{u'\left(c_t\right)}_{\substack{\text{Utility lost if you}\\ \text{invest "one" more}\\ \text{unit, i.e. marginal}\\ \text{cost of saving}}}=\underbrace{\beta u'\left(c_{t+1}\right)}_{\substack{\text{Utility increase}\\ \text{next period per}\\ \text{unit of increase in } c_{t+1}}}\cdot\underbrace{f'\left(k_{t+1}\right)}_{\substack{\text{Return on the}\\ \text{invested unit: by how}\\ \text{many units next period's}\\ c \text{ can increase}}}.$$

Thus, because of the concavity of u, equalizing the marginal cost of saving to the marginal benefit of saving is a condition for an optimum.

How do the primitives affect savings behavior? We can identify three component determinants of saving: the concavity of utility, the discounting, and the return to saving. Their effects are described in turn.

(i) Consumption "smoothing": if the utility function is *strictly* concave, the individual prefers a smooth consumption stream.

Example: Suppose that technology is linear, i.e. $f(k) = Rk$, and that $R\beta = 1$. Then

$$\beta f'(k_{t+1}) = \beta R = 1 \ \Rightarrow u'(c_t) = u'(c_{t+1}) \underbrace{\qquad\qquad\Rightarrow\qquad\qquad}_{\text{if } u \text{ is strictly concave}} c_t = c_{t+1}.$$

(ii) Impatience: via β, we see that a low β (a low discount factor, or a high discount rate $\frac{1}{\beta} - 1$) will tend to be associated with low c_{t+1}'s and high c_t's.

(iii) The return to savings: $f'(k_{t+1})$ clearly also affects behavior, but its effect on consumption cannot be signed unless we make more specific assumptions. Moreover, k_{t+1} is endogenous, so when f' nontrivially depends on it, we cannot vary the return independently. The case when f' is a constant, such as in the Ak growth model, is more convenient. We will return to it below.

To gain some more detailed understanding of the determinants of savings, let us study some examples.

Example 4.1 ***Logarithmic utility.*** *Let the utility index be*

$$u(c) = \log c,$$

and the production technology be represented by the function

$$f(k) = Rk.$$

Notice that this amounts to a linear function with exogenous marginal return R on investment.

The Euler equation becomes:

$$\begin{aligned} u'(c_t) &= \beta u'(c_{t+1}) \underbrace{f'(k_{t+1})}_{R} \\ \frac{1}{c_t} &= \frac{\beta R}{c_{t+1}}, \end{aligned}$$

and so

$$c_{t+1} = \beta R c_t. \tag{4.1}$$

The optimal path has consumption growing at the rate βR, and it is constant between any two periods. From the resource constraint (recall that it binds):

$$\begin{gathered} c_0 + k_1 = Rk_0 \\ c_1 + k_2 = Rk_1 \\ \vdots \\ c_T + k_{T+1} = Rk_T \\ k_{T+1} = 0. \end{gathered}$$

With repeated substitutions, we obtain the "consolidated" or "intertemporal" budget constraint:

$$c_0 + \frac{1}{R}c_1 + \frac{1}{R^2}c_2 + ... + \frac{1}{R^T}c_T = Rk_0.$$

The left-hand side is the present value of the consumption stream, and the right hand side is the present value of income. Using the optimal consumption growth rule $c_{t+1} = \beta R c_t$,

$$\begin{aligned} c_0 + \frac{1}{R}\beta R c_0 + \frac{1}{R^2}\beta^2 R^2 c_0 + ... + \frac{1}{R^T}\beta^T R^T c_0 &= Rk_0 \\ c_0 \left[1 + \beta + \beta^2 + ... + \beta^T\right] &= Rk_0. \end{aligned}$$

This implies

$$c_0 = \frac{Rk_0}{1 + \beta + \beta^2 + ... + \beta^T}.$$

We are now able to study the effects of changes in the marginal return on savings, R, on the consumer's behavior. An increase in R will cause a rise in consumption ***in all periods****. Crucial to this result is the chosen form for the utility function. Logarithmic utility has the property that income and substitution effects, when they go in opposite directions, exactly offset each other. Changes in R have two components: a change in relative prices (of consumption in different periods) and a change in present-value income: Rk_0. With logarithmic utility, a relative price change between two goods will make the consumption of the favored good go up whereas the consumption of other good will remain at the same level. The unfavored good will not be consumed in a lower amount since there is a positive income effect of the other good being cheaper, and that effect will be spread over both goods. Thus, the period 0 good will be unfavored in our example (since all other goods have lower price relative to good 0 if R goes up), and its consumption level will not decrease. The consumption of good 0 will in fact increase because total present-value income is multiplicative in R.*

Next assume that the sequence of interest rates is not constant, but that instead we have $\{R_t\}_{t=0}^T$ with R_t different at each t. The consolidated budget constraint now reads:

$$c_0 + \frac{1}{R_1}c_1 + \frac{1}{R_1 R_2}c_2 + \frac{1}{R_1 R_2 R_3}c_3 + ... + \frac{1}{R_1 ... R_T}c_T = k_0 R_0.$$

Plugging in the optimal path $c_{t+1} = \beta R_{t+1} c_t$, analogous to (4.1), one obtains

$$c_0 \left[1 + \beta + \beta^2 + ... + \beta^T\right] = k_0 R_0,$$

from which

$$\begin{aligned} c_0 &= \frac{k_0 R_0}{1+\beta+\beta^2+...+\beta^T} \\ c_1 &= \frac{k_0 R_0 R_1 \beta}{1+\beta+\beta^2+...+\beta^T} \\ &\vdots \\ c_t &= \frac{k_0 R_0 ... R_t \beta^t}{1+\beta+\beta^2+...+\beta^T}. \end{aligned}$$

Now note the following comparative statics:

$$\begin{aligned} R_t \uparrow \ &\Rightarrow c_0, c_1, ..., c_{t-1} \text{ are unaffected} \\ &\Rightarrow \text{ savings at } 0, \ ..., \ t-1 \text{ are unaffected.} \end{aligned}$$

In the logarithmic utility case, if the return between t and $t+1$ changes, consumption and savings remain unaltered until $t-1$!

Example 4.2 ***A slightly more general utility function.*** *Let us introduce the most commonly used additively separable utility function in macroeconomics: the CIES (constant intertemporal elasticity of substitution) function:*

$$u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}.$$

This function has as special cases:

$\sigma = 0$	*linear utility,*
$\sigma > 0$	*strictly concave utility,*
$\sigma = 1$	*logarithmic utility,*
$\sigma = \infty$	*not possible, but this is usually referred to as Leontief utility function.*

Let us define the intertemporal elasticity of substitution (IES):

$$IES \equiv \frac{\frac{d\left(\frac{c_{t+k}}{c_t}\right)}{\frac{c_{t+k}}{c_t}}}{\frac{dR_{t,t+k}}{R_{t,t+k}}}.$$

We will show that all the special cases of the CIES function have constant intertemporal elasticity of substitution equal to $\frac{1}{\sigma}$. We begin with the Euler equation:

$$u'(c_t) = \beta u'(c_{t+1}) R_{t+1}.$$

Replacing repeatedly, we have

$$\begin{aligned} u'(c_t) &= \beta^k u'(c_{t+k}) \underbrace{R_{t+1}R_{t+2}...R_{t+k}}_{\equiv R_{t,t+k}} \\ u'(c) &= c^{-\sigma} \Rightarrow c_t^{-\sigma} = \beta^k c_{t+k}^{-\sigma} R_{t,t+k} \\ \frac{c_{t+k}}{c_t} &= \left(\beta^k\right)^{\frac{1}{\sigma}} \left(R_{t,t+k}\right)^{\frac{1}{\sigma}}. \end{aligned}$$

This means that our elasticity measure becomes

$$\frac{\frac{d\left(\frac{c_{t+k}}{c_t}\right)}{\frac{c_{t+k}}{c_t}}}{\frac{dR_{t,t+k}}{R_{t,t+k}}} = \frac{d\log\frac{c_{t+k}}{c_t}}{d\log R_{t,t+k}} = \frac{1}{\sigma}.$$

When $\sigma = 1$, expenditure shares do not change: this is the logarithmic case. When $\sigma > 1$, an increase in $R_{t,t+k}$ would lead c_t to go up and savings to go down: the income effect, leading to smoothing across all goods, is larger than substitution effect. Finally, when $\sigma < 1$, the substitution effect is stronger: savings go up whenever $R_{t,t+k}$ goes up. When $\sigma = 0$, the elasticity is infinite and savings respond discontinuously to $R_{t,t+k}$.

4.1.2 Infinite horizon

Why should macroeconomists study the case of an infinite time horizon? There are at least two reasons:

1. *Altruism*: People do not live forever, but they may care about their offspring. Let $u(c_t)$ denote the utility flow to generation t. We can then interpret β^t as the weight an individual attaches to the utility enjoyed by his descendants t generations down the family tree. His total joy is given by $\sum_{t=0}^{\infty} \beta^t u(c_t)$. A $\beta < 1$ thus implies that the individual cares more about himself than about his descendants.

   If generations were overlapping the utility function would look similar:

   $$\sum_{t=0}^{\infty} \beta^t \underbrace{\left[u(c_{yt}) + \delta u(c_{ot})\right]}_{\text{utility flow to generation } t}.$$

   The existence of bequests indicates that there is altruism. However, bequests can also be of an entirely selfish, precautionary nature: when the life-time is unknown, as it is in practice, bequests would then be accidental and simply reflect the remaining buffer the individual kept for the possible remainder of his life. An argument for why bequests may not be entirely accidental is that annuity markets are not used very much. Annuity markets allow you to effectively insure against living "too long", and would thus make bequests disappear: all your wealth would be put into annuities and disappear upon death.

It is important to point out that the time horizon for an individual only becomes truly infinite if the altruism takes the form of caring about the utility of the descendants. If, instead, utility is derived from the act of giving itself, without reference to how the gift influences others' welfare, the individual's problem again becomes finite. Thus, if I live for one period and care about how much I give, my utility function might be $u(c)+v(b)$, where v measures how much I enjoy giving bequests, b. Although b subsequently shows up in another agent's budget and influences his choices and welfare, those effects are irrelevant for the decision of the present agent, and we have a simple static framework. This model is usually referred to as the "warm glow" model (the giver feels a warm glow from giving).

For a variation, think of an individual (or a dynasty) that, if still alive, each period dies with probability π. Its expected lifetime utility from a consumption stream $\{c_t\}_{t=0}^{\infty}$ is then given by

$$\sum_{t=0}^{\infty} \beta^t \pi^t u\left(c_t\right).$$

This framework - the "perpetual-youth" model, or, perhaps better, the "sudden-death" model - is sometimes used in applied contexts. Analytically, it looks like the infinite-life model, only with the difference that the discount factor is $\beta\pi$. These models are thus the same on the individual level. On the aggregate level, they are not, since the sudden-death model carries with it the assumption that a deceased dynasty is replaced with a new one: it is, formally speaking, an overlapping-generations model (see more on this below), and as such it is different in certain key respects.

Finally, one can also study explicit games between players of different generations. We may assume that parents care about their children, that sons care about their parents as well, and that each of their activities is in part motivated by this altruism, leading to intergenerational gifts as well as bequests. Since such models lead us into game theory rather quickly, and therefore typically to more complicated characterizations, we will assume that altruism is unidirectional.

2. *Simplicity*: Many macroeconomic models with a long time horizon tend to show very similar results to infinite-horizon models if the horizon is long enough. Infinite-horizon models are stationary in nature - the remaining time horizon does not change as we move forward in time - and their characterization can therefore often be obtained more easily than when the time horizon changes over time.

   The similarity in results between long- and infinite-horizon setups is is not present in all models in economics. For example, in the dynamic game theory the Folk Theorem means that the extension from a long (but finite) to an infinite horizon introduces a qualitative change in the model results. The typical example of this "discontinuity at infinity" is the prisoner's dilemma repeated a finite number of times, leading to a unique, non-cooperative outcome, versus the same game repeated an infinite number of times, leading to a large set of equilibria.

Models with an infinite time horizon demand more advanced mathematical tools. Consumers in our models are now choosing infinite sequences. These are no longer elements of Euclidean space $\Re^n$, which was used for our finite-horizon case. A basic question is when solutions to a given problem exist. Suppose we are seeking to maximize a function $U(x)$, $x \in S$. If $U(\cdot)$ is a continuous function, then we can invoke Weierstrass's theorem provided that the set S meets the appropriate conditions: S needs to be nonempty and compact. For $S \subset \Re^n$, compactness simply means closedness and boundedness. In the case of finite horizon, recall that x was a consumption vector of the form $(c_1, ..., c_T)$ from a subset S of $\Re^T$. In these cases, it was usually easy to check compactness. But now we have to deal with larger spaces; we are dealing with infinite-dimensional sequences $\{k_t\}_{t=0}^{\infty}$. Several issues arise. How do we define continuity in this setup? What is an open set? What does compactness mean? We will not answer these questions here, but we will bring up some specific examples of situations when maximization problems are ill-defined, that is, when they have no solution.

Examples where utility may be unbounded

Continuity of the objective requires boundedness. When will U be bounded? If two consumption streams yield "infinite" utility, it is not clear how to compare them. The device chosen to represent preference rankings over consumption streams is thus failing. But is it possible to get unbounded utility? How can we avoid this pitfall?

Utility may become unbounded for many reasons. Although these reasons interact, let us consider each one independently.

<u>Preference requirements</u>

Consider a plan specifying equal amounts of consumption goods for each period, throughout eternity:

$$\{c_t\}_{t=0}^{\infty} = \{\overline{c}\}_{t=0}^{\infty}.$$

Then the value of this consumption stream according to the chosen time-separable utility function representation is computed by:

$$U = \sum_{t=0}^{\infty} \beta^t u(c_t) = \sum_{t=0}^{\infty} \beta^t u(\overline{c}).$$

What is a necessary condition for U to take on a finite value in this case? The answer is $\beta < 1$: under this parameter specification, the series $\sum_{t=0}^{\infty} \beta^t u(\overline{c})$ is convergent, and has a finite limit. If $u(\cdot)$ has the CIES parametric form, then the answer to the question of convergence will involve not only β, but also σ.

Alternatively, consider a constantly increasing consumption stream:

$$\{c_t\}_{t=0}^{\infty} = \left\{c_0 (1+\gamma)^t\right\}_{t=0}^{\infty}.$$

Is $U = \sum_{t=0}^{\infty} \beta^t u(c_t) = \sum_{t=0}^{\infty} \beta^t u\left(c_0 (1+\gamma)^t\right)$ bounded? Notice that the argument in the instantaneous utility index $u(\cdot)$ is increasing without bound, while for $\beta < 1$ β^t is decreasing to 0. This seems to hint that the key to having a convergent series this time lies in the form of $u(\cdot)$ and in how it "processes" the increase in the value of its argument. In the case of a CIES utility representation, the relationship between β, σ, and γ is thus the key to boundedness. In particular, boundedness requires $\beta (1+\gamma)^{1-\sigma} < 1$.

Two other issues are involved in the question of boundedness of utility. One is technological, and the other may be called institutional.

Technological considerations

Technological restrictions are obviously necessary in some cases, as illustrated indirectly above. Let the technological constraints facing the consumer be represented by the budget constraint:

$$\begin{aligned} c_t + k_{t+1} &= Rk_t \\ k_t &\geq 0. \end{aligned}$$

This constraint needs to hold for all time periods t (this is just the "Ak" case already mentioned). This implies that consumption can grow by (at most) a rate of R. A given rate R may thus be so high that it leads to unbounded utility, as shown above.

Institutional framework

Some things simply cannot happen in an organized society. One of these is so dear to analysts modelling infinite-horizon economies that it has a name of its own. It expresses the fact that if an individual announces that he plans to borrow and never pay back, then he will not be able to find a lender. The requirement that "no Ponzi games are allowed" therefore represents this institutional assumption, and it sometimes needs to be added formally to the budget constraints of a consumer.

To see why this condition is necessary, consider a candidate solution to consumer's maximization problem $\{c_t^*\}_{t=0}^{\infty}$, and let $c_t^* \leq \bar{c}\ \forall t$; i.e., the consumption is bounded for every t. Suppose we endow a consumer with a given initial amount of net assets, a_0. These represent (real) claims against other agents. The constraint set is assumed to be

$$c_t + a_{t+1} = Ra_t, \forall t \geq 0.$$

Here $a_t < 0$ represents borrowing by the agent. Absent no-Ponzi-game condition, the agent could improve on $\{c_t^*\}_{t=0}^{\infty}$ as follows:

1. Put $\tilde{c}_0 = c_0^* + 1$, thus making $\tilde{a}_1 = a_1^* - 1$.

2. For every $t \geq 1$ leave $\tilde{c}_t = c_t^*$ by setting $\tilde{a}_{t+1} = a_{t+1}^* - R^t$.

With strictly monotone utility function, the agent will be strictly better off under this alternative consumption allocation, and it also satisfies budget constraint period-by-period.

Because this sort of improvement is possible for *any* candidate solution, the maximum of the lifetime utility will not exist.

However, observe that there is something wrong with the suggested improvement, as the agent's debt is growing without bound at rate R, and it is never repaid. This situation when the agent never repays his debt (or, equivalently, postpones repayment indefinitely) is ruled out by imposing the no-Ponzi-game (nPg) condition, by explicitly adding the restriction that:

$$\lim_{t\to\infty}\frac{a_t}{R^t}\geq 0.$$

Intuitively, this means that in present-value terms, the agent cannot engage in borrowing and lending so that his "terminal asset holdings" are negative, since this means that he would borrow and not pay back.

Can we use the nPg condition to simplify, or "consolidate", the sequence of budget constraints? By repeatedly replacing T times, we obtain

$$\sum_{t=0}^{T} c_t\frac{1}{R^t}+\frac{a_{T+1}}{R^T}\leq a_0R.$$

By the nPg condition, we then have

$$\begin{aligned}\lim_{T\to\infty}\left(\sum_{t=0}^{T} c_t\frac{1}{R^t}+\frac{a_{T+1}}{R^T}\right) &= \lim_{T\to\infty}\sum_{t=0}^{T} c_t\frac{1}{R^t}+\lim_{T\to\infty}\left(\frac{a_{T+1}}{R^T}\right)\\ &\equiv \sum_{t=0}^{\infty} c_t\frac{1}{R^t}+\lim_{T\to\infty}\left(\frac{a_{T+1}}{R^T}\right),\end{aligned}$$

and since the inequality is valid for every T, and we assume nPg condition to hold,

$$\sum_{t=0}^{\infty} c_t\frac{1}{R^t}\leq a_0R.$$

This is the consolidated budget constraint. In practice, we will often use a version of nPg with equality.

Example 4.3 *We will now consider a simple example that will illustrate the use of nPg condition in infinite-horizon optimization. Let the period utility of the agent* $u(c)=\log c$*, and suppose that there is one asset in the economy that pays a (net) interest rate of* r*. Assume also that the agent lives forever. Then, his optimization problem is:*

$$\begin{aligned}\max_{\{c_t,a_{t+1}\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty}\beta^t\log c_t\\ s.t. \quad & c_t+a_{t+1}=a_t(1+r)\,,\forall t\geq 0\\ & a_0 \text{ given}\\ & nPg \text{ condition.}\end{aligned}$$

To solve this problem, replace the period budget constraints with a consolidated one as we have done before. The consolidated budget constraint reads

$$\sum_{t=0}^{\infty} c_t \left(\frac{1}{1+r}\right)^t = a_0 (1+r).$$

With this simplification the first-order conditions are

$$\beta^t \frac{1}{c_t} = \lambda \left(\frac{1}{1+r}\right)^t, \forall t \geq 0,$$

where λ is the Lagrange multiplier associated with the consolidated budget constraint. From the first-order conditions it follows that

$$c_t = [\beta (1+r)]^t c_0, \forall t \geq 1.$$

Substituting this expression into the consolidated budget constraint, we obtain

$$\begin{aligned}
\sum_{t=0}^{\infty} \beta^t (1+r)^t \frac{1}{(1+r)^t} c_0 &= a_0 (1+r) \\
c_0 \sum_{t=0}^{\infty} \beta^t &= a_0 (1+r).
\end{aligned}$$

From here, $c_0 = a_0 (1-\beta)(1+r)$, and consumption in the periods $t \geq 1$ can be recovered from $c_t = [\beta (1+r)]^t c_0$.

Sufficient conditions

Maximization of utility under an infinite horizon will mostly involve the same mathematical techniques as in the finite-horizon case. In particular, we will make use of (Kuhn-Tucker) first-order conditions: barring corner constraints, we will choose a path such that the marginal effect of any choice variable on utility is zero. In particular, consider the sequences that the consumer chooses for his consumption and accumulation of capital. The first-order conditions will then lead to an Euler equation, which is defined for any path for capital beginning with an initial value k_0. In the case of finite time horizon it did not make sense for the agent to invest in the final period T, since no utility would be enjoyed from consuming goods at time $T+1$ when the economy is inactive. This final zero capital condition was key to determining the optimal path of capital: it provided us with a terminal condition for a difference equation system. In the case of infinite time horizon there is no such final T: the economy will continue forever. Therefore, the difference equation that characterizes the first-order condition may have an infinite number of solutions. We will need some other way of pinning down the consumer's choice, and it turns out that the missing condition is analogous to the requirement that the capital stock be zero at $T+1$, for else the consumer could increase his utility.

The missing condition, which we will now discuss in detail, is called the *transversality* condition. It is, typically, a necessary condition for an optimum, and it expresses the following simple idea: it cannot be optimal for the consumer to choose a capital sequence such that, in present-value utility terms, the shadow value of k_t remains positive as t goes to infinity. This could not be optimal because it would represent saving too much: a reduction in saving would still be feasible and would increase utility.

We will not prove the necessity of the transversality condition here. We will, however, provide a sufficiency condition. Suppose that we have a convex maximization problem (utility is concave and the constraint set convex) and a sequence $\{k_{t+1}\}_{t=1}^{\infty}$ satisfying the Kuhn-Tucker first-order conditions for a given k_0. Is $\{k_{t+1}\}_{t=1}^{\infty}$ a maximum? We did not formally prove a similar proposition in the finite-horizon case (we merely referred to math texts), but we will here, and the proof can also be used for finite-horizon setups.

Sequences satisfying the Euler equations that do not maximize the programming problem come up quite often. We would like to have a systematic way of distinguishing between maxima and other critical points (in $\Re^{\infty}$) that are not the solution we are looking for. Fortunately, the transversality condition helps us here: if a sequence $\{k_{t+1}\}_{t=1}^{\infty}$ satisfies both the Euler equations and the transversality condition, then it maximizes the objective function. Formally, we have the following:

Proposition 4.4 *Consider the programming problem*

$$\begin{aligned} \max_{\{k_{t+1}\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t F\left(k_t,\ k_{t+1}\right) \\ s.t. \quad & k_{t+1} \geq 0\ \forall t. \end{aligned}$$

(An example is $F(x,y) = u\left[f\left(x\right) - y\right]$.)

If $\left\{k_{t+1}^*\right\}_{t=0}^{\infty}$, $\{\mu_t^*\}_{t=0}^{\infty}$ satisfy

(i) $k_{t+1}^* \geq 0\ \forall t$

(ii) Euler Equation: $F_2\left(k_t^*,\ k_{t+1}^*\right) + \beta F_1\left(k_{t+1}^*,\ k_{t+2}^*\right) + \mu_t^* = 0\ \forall t$

(iii) $\mu_t^* \geq 0$, $\mu_t^* k_{t+1}^* = 0\ \forall t$

(iv) $\lim\limits_{t\to\infty} \beta^t F_1\left(k_t^*,\ k_{t+1}^*\right) k_t^* = 0$

and $F\left(x,y\right)$ is concave in $\left(x,y\right)$ and increasing in its first argument, then $\left\{k_{t+1}^*\right\}_{t=0}^{\infty}$ maximizes the objective.

Proof. Consider any alternative feasible sequence $\mathbf{k} \equiv \{k_{t+1}\}_{t=0}^{\infty}$. Feasibility is tantamount to $k_{t+1} \geq 0\ \forall t$. We want to show that for any such sequence,

$$\lim_{T\to\infty} \sum_{t=0}^{T} \beta^t \left[F\left(k_t^*,\ k_{t+1}^*\right) - F\left(k_t,\ k_{t+1}\right)\right] \geq 0.$$

Define

$$A_T(\mathbf{k}) \equiv \sum_{t=0}^{T} \beta^t \left[F\left(k_t^*,\ k_{t+1}^*\right) - F\left(k_t,\ k_{t+1}\right)\right].$$

We will to show that, as T goes to infinity, $A_T(\mathbf{k})$ is bounded below by zero.

By concavity of F,

$$A_T(\mathbf{k}) \geq \sum_{t=0}^{T} \beta^t \left[F_1\left(k_t^*,\ k_{t+1}^*\right)\left(k_t^* - k_t\right) + F_2\left(k_t^*,\ k_{t+1}^*\right)\left(k_{t+1}^* - k_{t+1}\right)\right].$$

Now notice that for each t, k_{t+1} shows up twice in the summation. Hence we can rearrange the expression to read

$$\begin{aligned} A_T(\mathbf{k}) \ \geq\ & \sum_{t=0}^{T-1} \beta^t \left\{\left(k_{t+1}^* - k_{t+1}\right)\left[F_2\left(k_t^*,\ k_{t+1}^*\right) + \beta F_1\left(k_{t+1}^*,\ k_{t+2}^*\right)\right]\right\} + \\ & + F_1\left(k_0^*,\ k_1^*\right)\left(k_0^* - k_0\right) + \beta^T F_2\left(k_T^*,\ k_{T+1}^*\right)\left(k_{T+1}^* - k_{T+1}\right). \end{aligned}$$

Some information contained in the first-order conditions will now be useful:

$$F_2\left(k_t^*,\ k_{t+1}^*\right) + \beta F_1\left(k_{t+1}^*,\ k_{t+2}^*\right) = -\mu_t^*,$$

together with $k_0^* - k_0 = 0$ (k_0 can only take on one feasible value), allows us to derive

$$A_T(\mathbf{k}) \geq \sum_{t=0}^{T-1} \beta^t \mu_t^*\left(k_{t+1} - k_{t+1}^*\right) + \beta^T F_2\left(k_T^*,\ k_{T+1}^*\right)\left(k_{T+1}^* - k_{T+1}\right).$$

Next, we use the complementary slackness conditions and the implication of the Kuhn-Tucker conditions that

$$\mu_t^* k_{t+1} \geq 0$$

to conclude that $\mu_t^*\left(k_{t+1} - k_{t+1}^*\right) \geq 0$. In addition, $F_2\left(k_T^*,\ k_{T+1}^*\right) = -\beta F_1\left(k_{T+1}^*,\ k_{T+2}^*\right) - \mu_T^*$, so we obtain

$$A_T(\mathbf{k}) \geq \sum_{t=0}^{T} \beta^t \mu_t^*\left(k_{t+1} - k_{t+1}^*\right) + \beta^T\left[\beta F_1\left(k_{T+1}^*,\ k_{T+2}^*\right) + \mu_T^*\right]\left(k_{T+1} - k_{T+1}^*\right).$$

Since we know that $\mu_t^*\left(k_{t+1} - k_{t+1}^*\right) \geq 0$, the value of the summation will not increase if we suppress nonnegative terms:

$$A_T(\mathbf{k}) \geq \beta^{T+1} F_1\left(k_{T+1}^*,\ k_{T+2}^*\right)\left(k_{T+1} - k_{T+1}^*\right) \geq -\beta^{T+1} F_1\left(k_{T+1}^*,\ k_{T+2}^*\right) k_{T+1}^*.$$

In the finite horizon case, k_{T+1}^* would have been the level of capital left out for the day after the (perfectly foreseen) end of the world; a requirement for an optimum in that case is clearly $k_{T+1}^* = 0$. In present-value utility terms, one might alternatively require $k_{T+1}^* \beta^T \lambda_T^* = 0$, where $\beta^t \lambda_t^*$ is the present-value utility evaluation of an additional unit of resources in period t.

As T goes to infinity, the right-hand side of the last inequality goes to zero by the transversality condition. That is, we have shown that the utility implied by the candidate path must be higher than that implied by the alternative. ■

The transversality condition can be given this interpretation: $F_1(k_t, k_{t+1})$ is the marginal addition of utils in period t from increasing capital in that period, so the transversality condition simply says that the value (discounted into present-value utils) of each additional unit of capital at infinity times the actual amount of capital has to be zero. If this requirement were not met (we are now, incidentally, making a heuristic argument for necessity), it would pay for the consumer to modify such a capital path and increase consumption for an overall increase in utility without violating feasibility.[1]

The no-Ponzi-game and the transversality conditions play very similar roles in dynamic optimization in a purely mechanical sense (at least if the nPg condition is interpreted with equality). In fact, they can typically be shown to be the same condition, if one also assumes that the first-order condition is satisfied. However, the two conditions are conceptually very different. The nPg condition is a restriction on the choices of the agent. In contrast, the transversality condition is a prescription how to behave optimally, *given* a choice set.

4.2 Dynamic programming

The models we are concerned with consist of a more or less involved dynamic optimization problem and a resulting optimal consumption plan that solves it. Our approach up to now has been to look for a sequence of real numbers $\left\{k_{t+1}^*\right\}_{t=0}^{\infty}$ that generates an optimal consumption plan. In principle, this involved searching for a solution to an infinite sequence of equations - a difference equation (the Euler equation). The search for a sequence is sometimes impractical, and not always intuitive. An alternative approach is often available, however, one which is useful conceptually as well as for computation (both analytical and, especially, numerical computation). It is called dynamic programming. We will now go over the basics of this approach. The focus will be on concepts, as opposed to on the mathematical aspects or on the formal proofs.

Key to dynamic programming is to think of dynamic decisions as being made not once and for all but recursively: time period by time period. The savings between t and $t+1$ are thus decided on at t, and not at 0. We will call a problem *stationary* whenever the structure of the choice problem that a decision maker faces is identical at every point in time. As an illustration, in the examples that we have seen so far, we posited a consumer placed at the beginning of time choosing his infinite future consumption stream given an initial capital stock k_0. As a result, out came a sequence of real numbers $\left\{k_{t+1}^*\right\}_{t=0}^{\infty}$ indicating the level of capital that the agent will choose to hold in each period. But once he has chosen a capital path, suppose that we let the consumer abide it for, say, T periods. At $t = T$ he will find

[1]This necessity argument clearly requires utility to be strictly increasing in capital.

then himself with the k_T^* decided on initially. If at that moment we told the consumer to forget about his initial plan and asked him to decide on his consumption stream again, from then onwards, using as new initial level of capital $k_0 = k_T^*$, what sequence of capital would he choose? If the problem is *stationary* then for any two periods $t \neq s$,

$$k_t = k_s \Rightarrow k_{t+j} = k_{s+j}$$

for all $j > 0$. That is, he would not change his mind if he could decide all over again.

This means that, if a problem is stationary, we can think of a function that, for every period t, assigns to each possible initial level of capital k_t an optimal level for next period's capital k_{t+1} (and therefore an optimal level of current period consumption): $k_{t+1} = g(k_t)$. Stationarity means that the function $g(\cdot)$ has no other argument than current capital. In particular, the function does not vary with time. We will refer to $g(\cdot)$ as the *decision rule.*

We have defined stationarity above in terms of decisions - in terms of properties of the solution to a dynamic problem. What types of dynamic problems are stationary? Intuitively, a dynamic problem is stationary if one can capture all relevant information for the decision maker in a way that does not involve time. In our neoclassical growth framework, with a finite horizon, time is important, and the problem is not stationary: it matters how many periods are left - the decision problem changes character as time passes. With an infinite time horizon, however, the remaining horizon is the same at each point in time. The only changing feature of the consumer's problem in the infinite-horizon neoclassical growth economy is his initial capital stock; hence, his decisions will not depend on anything but this capital stock. Whatever is the relevant information for a consumer solving a dynamic problem, we will refer to it as his *state variable*. So the state variable for the planner in the one-sector neoclassical growth context is the current capital stock.

The heuristic information above can be expressed more formally as follows. The simple mathematical idea that $\max_{x,y} f(x,y) = \max_y\{\max_x f(x,y)\}$ (if each of the max operators is well-defined) allows us to maximize "in steps": first over x, given y, and then the remainder (where we can think of x as a function of y) over y. If we do this over time, the idea would be to maximize over $\{k_{s+1}\}_{s=t}^{\infty}$ first by choice of $\{k_{s+1}\}_{s=t+1}^{\infty}$, conditional on k_{t+1}, and then to choose k_{t+1}. That is, we would choose savings at t, and later the rest. Let us denote by $V(k_t)$ the value of the optimal program from period t for an initial condition k_t:

$$V(k_t) \equiv \max_{\{k_{s+1}\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \beta^{s-t} F(k_s, k_{s+1}), \text{ s.t. } k_{s+1} \in \Gamma(k_s) \forall s \geq t,$$

where $\Gamma(k_t)$ represents the feasible choice set for k_{t+1} given k_t[2]. That is, V is an indirect utility function, with k_t representing the parameter governing the choices and resulting utility. Then using the maximization-by-steps idea, we can write

$$V(k_t) = \max_{k_{t+1} \in \Gamma(k_t)} \{F(k_t, k_{t+1}) + \max_{\{k_{s+1}\}_{s=t+1}^{\infty}} \sum_{s=t+1}^{\infty} \beta^{s-t} F(k_s, k_{s+1}) \text{ (s.t. } k_{s+1} \in \Gamma(k_s) \forall s \geq t+1)\},$$

[2]The one-sector growth model example would mean that $F(x,y) = u(f(x)-y)$ and that $\Gamma(x) = [0, f(x)]$ (the latter restricting consumption to be non-negative and capital to be non-negative).

which in turn can be rewritten as

$$\max_{k_{t+1}\in\Gamma(k_t)}\{F(k_t,k_{t+1})+\beta\max_{\{k_{s+1}\}_{s=t+1}^{\infty}}\{\sum_{s=t+1}^{\infty}\beta^{s-(t+1)}F(k_s,k_{s+1})\text{ (s.t. } k_{s+1}\in\Gamma(k_s)\forall s\geq t+1)\}\}.$$

But by definition of V this equals

$$\max_{k_{t+1}\in\Gamma(k_t)}\{F(k_t,k_{t+1})+\beta V(k_{t+1})\}.$$

So we have:

$$V(k_t)=\max_{k_{t+1}\in\Gamma(k_t)}\{F(k_t,k_{t+1})+\beta V(k_{t+1})\}.$$

This is the dynamic programming formulation. The derivation was completed for a given value of k_t on the left-hand side of the equation. On the right-hand side, however, we need to know V evaluated at any value for k_{t+1} in order to be able to perform the maximization. If, in other words, we find a V that, using k to denote current capital and k' next period's capital, satisfies

$$V(k)=\max_{k'\in\Gamma(k)}\{F(k,k')+\beta V(k')\} \tag{4.2}$$

for any value of k, then all the maximizations on the right-hand side are well-defined. This equation is called the Bellman equation, and it is a *functional equation*: the unknown is a function. We use the function g alluded to above to denote the arg max in the functional equation:

$$g(k)=\arg\max_{k'\in\Gamma(k)}\{F(k,k')+\beta V(k')\},$$

or the decision rule for k': $k'=g(k)$. This notation presumes that a maximum exists and is unique; otherwise, g would not be a well-defined function.

This is "close" to a formal derivation of the equivalence between the sequential formulation of the dynamic optimization and its recursive, Bellman formulation. What remains to be done mathematically is to make sure that all the operations above are well-defined. Mathematically, one would want to establish:

- If a function represents the value of solving the sequential problem (for any initial condition), then this function solves the dynamic programming equation (DPE).

- If a function solves the DPE, then it gives the value of the optimal program in the sequential formulation.

- If a sequence solves the sequential program, it can be expressed as a decision rule that solves the maximization problem associated with the DPE.

- If we have a decision rule for a DPE, it generates sequences that solve the sequential problem.

These four facts can be proved, under appropriate assumptions.[3] We omit discussion of details here.

One issue is useful to touch on before proceeding to the practical implementation of dynamic programming: since the maximization that needs to be done in the DPE is finite-dimensional, ordinary Kuhn-Tucker methods can be used, without reference to extra conditions, such as the transversality condition. How come we do not need a transversality condition here? The answer is subtle and mathematical in nature. In the statements and proofs of equivalence between the sequential and the recursive methods, it is necessary to impose conditions on the function V: not any function is allowed. Uniqueness of solutions to the DPE, for example, only follows by restricting V to lie in a restricted space of functions. This or other, related, restrictions play the role of ensuring that the transversality condition is met.

We will make use of some important results regarding dynamic programming. They are summarized in the following:

Facts

Suppose that F is continuously differentiable in its two arguments, that it is strictly increasing in its first argument (and decreasing in the second), strictly concave, and bounded. Suppose that Γ is a nonempty, compact-valued, monotone, and continuous correspondence with a convex graph. Finally, suppose that $\beta \in (0,1)$. Then

1. There exists a function $V(\cdot)$ that solves the Bellman equation. This solution is unique.

2. It is possible to find V by the following iterative process:

   i. Pick any initial V_0 function, for example $V_0(k) = 0 \ \forall k$.

   ii. Find V_{n+1}, for any value of k, by evaluating the right-hand side of (4.2) using V_n.

   The outcome of this process is a sequence of functions $\{V_j\}_{j=0}^{\infty}$ which converges to V.

3. V is strictly concave.

4. V is strictly increasing.

5. V is continuously differentiable.

6. Optimal behavior can be characterized by a function g, with $k' = g(k)$, that is increasing so long as F_2 is increasing in k.

The proof of the existence and uniqueness part follow by showing that the functional equation's right-hand side is a contraction mapping, and using the contraction mapping theorem. The algorithm for finding V also uses the contraction property. The assumptions

[3] See Stokey and Lucas (1989).

needed for these characterizations do not rely on properties of F other than its continuity and boundedness. That is, these results are quite general.

In order to prove that V is increasing, it is necessary to assume that F is increasing and that Γ is monotone. In order to show that V is (strictly) concave it is necessary to assume that F is (strictly) concave and that Γ has a convex graph. Both these results use the iterative algorithm. They essentially require showing that, if the initial guess on V, V_0, satisfies the required property (such as being increasing), then so is any subsequent V_n. These proofs are straightforward.

Differentiability of V requires F to be continuously differentiable and concave, and the proof is somewhat more involved. Finally, optimal policy is a function when F is strictly concave and Γ is convex-valued; under these assumptions, it is also easy to show, using the first-order condition in the maximization, that g is increasing. This condition reads

$$-F_2(k, k') = \beta V'(k').$$

The left-hand side of this equality is clearly increasing in k', since F is strictly concave, and the right-hand side is strictly decreasing in k', since V is strictly concave under the stated assumptions. Furthermore, since the right-hand side is independent of k but the left-hand side is decreasing in k, the optimal choice of k' is increasing in k.

The proofs of all these results can be found in Stokey and Lucas with Prescott (1989).

Connection with finite-horizon problems

Consider the finite-horizon problem

$$\begin{aligned}&\max_{\{c_t\}_{t=0}^{T}} \sum_{t=0}^{T} \beta^t u\left(c_t\right)\\ &\text{s.t. } k_{t+1} + c_t = F\left(k_t\right).\end{aligned}$$

Although we discussed how to solve this problem in the previous sections, dynamic programming offers us a new solution method. Let $V_n(k)$ denote the present value utility derived from having a current capital stock of k and behaving optimally, if there are n periods left until the end of the world. Then we can solve the problem recursively, or by backward induction, as follows. If there are no periods left, that is, if we are at $t = T$, then the present value of utility next period will be 0 no matter how much capital is chosen to be saved: $V_0(k) = 0 \ \forall k$. Then once he reaches $t = T$ the consumer will face the following problem:

$$V_1(k) = \max_{k'} \left\{u\left[f(k) - k'\right] + \beta V_0(k')\right\}.$$

Since $V_0(k') = 0$, this reduces to $V_1(k) = \max_{k'} \{u[f(k) - k']\}$. The solution is clearly $k' = 0$ (note that this is consistent with the result $k_{T+1} = 0$ that showed up in finite horizon problems when the formulation was sequential). As a result, the update is $V_1(k) = u[f(k)]$. We can iterate in the same fashion T times, all the way to V_{T+1}, by successively plugging in the updates V_n. This will yield the solution to our problem.

In this solution of the finite-horizon problem, we have obtained an interpretation of the iterative solution method for the infinite-horizon problem: the iterative solution is like solving a finite-horizon problem backwards, for an increasing time horizon. The statement that the limit function converges says that the value function of the infinite-horizon problem is the limit of the time-zero value functions of the finite-horizon problems, as the horizon increases to infinity. This also means that the behavior at time zero in a finite-horizon problem becomes increasingly similar to infinite-horizon behavior as the horizon increases.

Finally, notice that we used dynamic programming to describe how to solve a non-stationary problem. This may be confusing, as we stated early on that dynamic programming builds on stationarity. However, if time is viewed as a state variable, as we actually did view it now, the problem can be viewed as stationary. That is, if we increase the state variable from not just including k, but t as well (or the number of periods left), then dynamic programming can again be used.

Example 4.5 ***Solving a parametric dynamic programming problem.*** *In this example we will illustrate how to solve dynamic programming problem by finding a corresponding value function. Consider the following functional equation:*

$$\begin{aligned} V(k) = \; & \max_{c,\, k'} \left\{ \log c + \beta V(k') \right\} \\ & s.t.\; c = Ak^{\alpha} - k'. \end{aligned}$$

The budget constraint is written as an equality constraint because we know that preferences represented by the logarithmic utility function exhibit strict monotonicity - goods are always valuable, so they will not be thrown away by an optimizing decision maker. The production technology is represented by a Cobb-Douglas function, and there is full depreciation of the capital stock in every period:

$$\underbrace{F(k,1)}_{Ak^{\alpha}1^{1-\alpha}} + \underbrace{(1-\delta)}_{0}k.$$

A more compact expression can be derived by substitutions into the Bellman equation:

$$V(k) = \max_{k' \geq 0} \left\{ \log \left[Ak^{\alpha} - k'\right] + \beta V(k') \right\}.$$

We will solve the problem by iterating on the value function. The procedure will be similar to that of solving a T-problem backwards. We begin with an initial "guess" $V_0(k) = 0$, that is, a function that is zero-valued everywhere.

$$\begin{aligned} V_1(k) &= \max_{k' \geq 0} \left\{ \log \left[Ak^{\alpha} - k'\right] + \beta V_0(k') \right\} \\ &= \max_{k' \geq 0} \left\{ \log \left[Ak^{\alpha} - k'\right] + \beta \cdot 0 \right\} \\ & \max_{k' \geq 0} \left\{ \log \left[Ak^{\alpha} - k'\right] \right\}. \end{aligned}$$

This is maximized by taking $k' = 0$. Then

$$V_1(k) = \log A + \alpha \log k.$$

Going to the next step in the iteration,

$$\begin{aligned} V_2(k) &= \max_{k' \geq 0} \{\log [Ak^\alpha - k'] + \beta V_1(k')\} \\ &= \max_{k' \geq 0} \{\log [Ak^\alpha - k'] + \beta [\log A + \alpha \log k']\}. \end{aligned}$$

The first-order condition now reads

$$\frac{1}{Ak^\alpha - k'} = \frac{\beta\alpha}{k'} \Rightarrow k' = \frac{\alpha\beta Ak^\alpha}{1+\alpha\beta}.$$

We can interpret the resulting expression for k' as the rule that determines how much it would be optimal to save if we were at period $T-1$ in the finite horizon model. Substitution implies

$$\begin{aligned} V_2(k) &= \log\left[Ak^\alpha - \frac{\alpha\beta Ak^\alpha}{1+\alpha\beta}\right] + \beta\left[\log A + \alpha\log\frac{\alpha\beta Ak^\alpha}{1+\alpha\beta}\right] \\ &= \left(\alpha + \alpha^2\beta\right)\log k + \log\left(A - \frac{\alpha\beta A}{1+\alpha\beta}\right) + \beta\log A + \alpha\beta\log\frac{\alpha\beta A}{1+\alpha\beta}. \end{aligned}$$

We could now use $V_2(k)$ again in the algorithm to obtain a $V_3(k)$, and so on. We know by the characterizations above that this procedure would make the sequence of value functions converge to some $V^*(k)$. However, there is a more direct approach, using a pattern that appeared already in our iteration.

Let

$$a \equiv \log\left(A - \frac{\alpha\beta A}{1+\alpha\beta}\right) + \beta\log A + \alpha\beta\log\frac{\alpha\beta A}{1+\alpha\beta}$$

and

$$b \equiv \left(\alpha + \alpha^2\beta\right).$$

Then $V_2(k) = a + b\log k$. Recall that $V_1(k) = \log A + \alpha\log k$, i.e., in the second step what we did was plug in a function $V_1(k) = a_1 + b_1\log k$, and out came a function $V_2(k) = a_2 + b_2\log k$. This clearly suggests that if we continue using our iterative procedure, the outcomes $V_3(k)$, $V_4(k)$, ..., $V_n(k)$, will be of the form $V_n(k) = a_n + b_n\log k$ for all n. Therefore, we may already guess that the function to which this sequence is converging has to be of the form:

$$V(k) = a + b\log k.$$

So let us guess that the value function solving the Bellman has this form, and determine the corresponding parameters a, b:

$$V(k) = a + b\log k = \max_{k' \geq 0} \{\log(Ak^\alpha - k') + \beta(a + b\log k')\} \;\; \forall k.$$

Our task is to find the values of a and b such that this equality holds for all possible values of k. If we obtain these values, the functional equation will be solved.

The first-order condition reads:

$$\frac{1}{Ak^\alpha - k'} = \frac{\beta b}{k'} \;\Rightarrow k' = \frac{\beta b}{1+\beta b} A k^\alpha .$$

We can interpret $\frac{\beta b}{1+\beta b}$ *as a savings rate. Therefore, in this setup the optimal policy will be to save a constant fraction out of each period's income.*

Define

$$LHS \equiv a + b \log k$$

and

$$RHS \equiv \max_{k' \geq 0} \left\{ \log\left(Ak^\alpha - k'\right) + \beta\left(a + b\log k'\right)\right\}.$$

Plugging the expression for k' *into the RHS, we obtain*:

$$\begin{aligned}
RHS &= \log\left(Ak^\alpha - \frac{\beta b}{1+\beta b}Ak^\alpha\right) + a\beta + b\beta \log\left(\frac{\beta b}{1+\beta b}Ak^\alpha\right) \\
&= \log\left[\left(1 - \frac{\beta b}{1+\beta b}\right)Ak^\alpha\right] + a\beta + b\beta \log\left(\frac{\beta b}{1+\beta b}Ak^\alpha\right) \\
&= (1+b\beta)\log A + \log\left(\frac{1}{1+b\beta}\right) + a\beta + b\beta\log\left(\frac{\beta b}{1+\beta b}\right) + (\alpha + \alpha\beta b)\log k.
\end{aligned}$$

Setting LHS=RHS, we produce

$$\begin{cases}
a = (1+b\beta)\log A + \log\left(\dfrac{1}{1+b\beta}\right) + a\beta + b\beta\log\left(\dfrac{\beta b}{1+\beta b}\right) \\
\\
b = \alpha + \alpha\beta b,
\end{cases}$$

which amounts to two equations in two unknowns. The solutions will be

$$b = \frac{\alpha}{1-\alpha\beta}$$

and, using this finding,

$$a = \frac{1}{1-\beta}\left[(1+b\beta)\log A + b\beta\log(b\beta) - (1+b\beta)\log(1+b\beta)\right],$$

so that

$$a = \frac{1}{1-\beta}\frac{1}{1-\alpha\beta}\left[\log A + (1-\alpha\beta)\log(1-\alpha\beta) + \alpha\beta\log(\alpha\beta)\right].$$

Going back to the savings decision rule, we have:

$$\begin{aligned}
k' &= \frac{b\beta}{1+b\beta}Ak^\alpha \\
k' &= \alpha\beta A k^\alpha .
\end{aligned}$$

If we let y denote income, that is, $y \equiv Ak^{\alpha}$, then $k' = \alpha\beta y$. This means that the optimal solution to the path for consumption and capital is to save a constant fraction $\alpha\beta$ of income.

This setting, we have now shown, provides a microeconomic justification to a constant savings rate, like the one assumed by Solow. It is a very special setup however, one that is quite restrictive in terms of functional forms. Solow's assumption cannot be shown to hold generally.

We can visualize the dynamic behavior of capital as is shown in Figure 4.1.

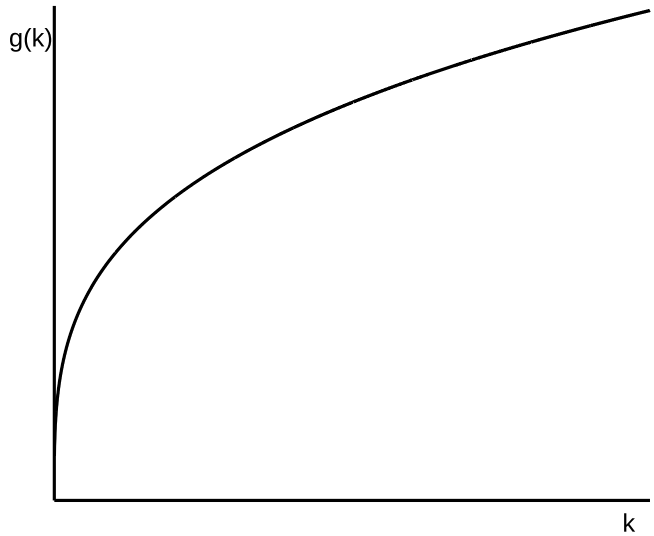

Figure 4.1: The decision rule in our parameterized model

Example 4.6 ***A more complex example***. *We will now look at a slightly different growth model and try to put it in recursive terms. Our new problem is:*

$$\begin{array}{ll} \max\limits_{\{c_t\}_{t=0}^{\infty}} & \sum\limits_{t=0}^{\infty} \beta^t u\left(c_t\right) \\ s.t. & c_t + i_t = F\left(k_t\right) \end{array}$$

and subject to the assumption is that capital depreciates fully in ***two*** *periods, and does not depreciate at all before that. Then the law of motion for capital, given a sequence of investment $\{i_t\}_{t=0}^{\infty}$ is given by:*

$$k_t = i_{t-1} + i_{t-2}.$$

Then $k = i_{-1} + i_{-2}$: there are two initial conditions i_{-1} and i_{-2}.

The recursive formulation for this problem is:

$$\begin{array}{rl} V\left(i_{-1},\ i_{-2}\right) = & \max\limits_{c,\ i} \left\{u(c) + V\left(i,\ i_{-1}\right)\right\} \\ s.t. & c = f\left(i_{-1} + i_{-2}\right) - i. \end{array}$$

Notice that there are two state variables in this problem. That is unavoidable here; there is no way of summarizing what one needs to know at a point in time with only one variable. For example, the total capital stock in the current period is not informative enough, because in order to know the capital stock next period we need to know how much of the current stock will disappear between this period and the next. Both i_{-1} and i_{-2} are natural state variables: they are predetermined, they affect outcomes and utility, and neither is redundant: the information they contain cannot be summarized in a simpler way.

4.3 The functional Euler equation

In the sequentially formulated maximization problem, the Euler equation turned out to be a crucial part of characterizing the solution. With the recursive strategy, an Euler equation can be derived as well. Consider again

$$V(k) = \max_{k' \in \Gamma(k)} \left\{ F(k,\ k') + \beta V(k') \right\}.$$

As already pointed out, under suitable assumptions, this problem will result in a function $k' = g(k)$ that we call decision rule, or policy function. By definition, then, we have

$$V(k) = F(k,\ g(k)) + \beta V[g(k)]. \tag{4.3}$$

Moreover, $g(k)$ satisfies the first-order condition

$$F_2(k,\ k') + \beta V'(k') = 0,$$

assuming an interior solution. Evaluating at the optimum, i.e., at $k' = g(k)$, we have

$$F_2(k,\ g(k)) + \beta V'(g(k)) = 0.$$

This equation governs the intertemporal tradeoff. One problem in our characterization is that $V'(\cdot)$ is not known: in the recursive strategy, it is part of what we are searching for. However, although it is not possible in general to write $V(\cdot)$ in terms of primitives, one can find its derivative. Using the equation (4.3) above, one can differentiate both sides with respect to k, since the equation holds for all k and, again under some assumptions stated earlier, is differentiable. We obtain

$$V'(k) = F_1[k,\ g(k)] + \underbrace{g'(k)\left\{F_2[k,\ g(k)] + \beta V'[g(k)]\right\}}_{\text{indirect effect through optimal choice of } k'}.$$

From the first-order condition, this reduces to

$$V'(k) = F_1[k,\ g(k)],$$

which again holds for all values of k. The indirect effect thus disappears: this is an application of a general result known as the envelope theorem.

Updating, we know that $V'[g(k)] = F_1[g(k),\ g(g(k))]$ also has to hold. The first order condition can now be rewritten as follows:

$$F_2[k,\ g(k)] + \beta F_1[g(k),\ g(g(k))] = 0\ \forall k. \tag{4.4}$$

This is the Euler equation stated as a functional equation: it does not contain the unknowns k_t, k_{t+1}, and k_{t+2}. Recall our previous Euler equation formulation

$$F_2[k_t,\ k_{t+1}] + \beta F_1[k_{t+1},\ k_{t+2}] = 0, \forall t,$$

where the unknown was the sequence $\{k_t\}_{t=1}^{\infty}$. Now instead, the unknown is the function g. That is, under the recursive formulation, the Euler Equation turned into a functional equation.

The previous discussion suggests that a third way of searching for a solution to the dynamic problem is to consider the functional Euler equation, and solve it for the function g. We have previously seen that we can (i) look for sequences solving a nonlinear difference equation plus a transversality condition; or (ii) we can solve a Bellman (functional) equation for a value function.

The functional Euler equation approach is, in some sense, somewhere in between the two previous approaches. It is based on an equation expressing an intertemporal tradeoff, but it applies more structure than our previous Euler equation. There, a transversality condition needed to be invoked in order to find a solution. Here, we can see that the recursive approach provides some extra structure: it tells us that the optimal sequence of capital stocks needs to be connected using a stationary function.

One problem is that the functional Euler equation does not in general have a unique solution for g. It might, for example, have two solutions. This multiplicity is less severe, however, than the multiplicity in a second-order difference equation without a transversality condition: there, there are infinitely many solutions.

The functional Euler equation approach is often used in practice in solving dynamic problems numerically. We will return to this equation below.

Example 4.7 *In this example we will apply functional Euler equation described above to the model given in Example 4.5. First, we need to translate the model into "V-F language". With full depreciation and strictly monotone utility function, the function $F(\cdot,\cdot)$ has the form*

$$F(k, k') = u(f(k) - g(k)).$$

Then, the respective derivatives are:

$$\begin{aligned} F_1(k, k') &= u'(f(k) - k')\, f'(k) \\ F_2(k, k') &= -u'(f(k) - k'). \end{aligned}$$

In the particular parametric example, (4.4) becomes:

$$\frac{1}{Ak^{\alpha} - g(k)} - \frac{\beta\alpha A\,(g(k))^{\alpha-1}}{A\,(g(k))^{\alpha} - g(g(k))} = 0, \forall k.$$

This is a functional equation in $g(k)$. Guess that $g(k) = sAk^{\alpha}$, i.e. the savings are a constant fraction of output. Substituting this guess into functional Euler equation delivers:

$$\frac{1}{(1-s)\,Ak^{\alpha}} = \frac{\alpha\beta A\,(sAk^{\alpha})^{\alpha-1}}{A\,(sAk^{\alpha})^{\alpha} - sA\,(sAk^{\alpha})^{\alpha}}.$$

As can be seen, k cancels out, and the remaining equation can be solved for s. Collecting terms and factoring out s, we get

$$s = \alpha\beta.$$

This is exactly the answer that we got in Example 4.5.

4.4 References

Stokey, Nancy L., and Robert E. Lucas, "Recursive Methods in Economic Dynamics", Harvard University Press, 1989.

Chapter 5

Competitive equilibrium in dynamic models

It is now time to leave pure maximization setups where there is a planner making all decisions and move on to market economies. What *economic arrangement*, or what *allocation mechanism*, will be used in the model economy to talk about decentralized, or at least less centralized, behavior? Of course, different physical environments may call for different arrangements. Although many argue that the modern market economy is not well described by well-functioning markets due to the presence of various frictions (incomplete information, externalities, market power, and so on), it still seems a good idea to build the frictionless economy first, and use it as a benchmark from which extensions can be systematically built and evaluated. For a frictionless economy, competitive equilibrium analysis therefore seems suitable.

One issue is what the population structure will be. We will first look at the infinite-horizon (dynastic) setup. The generalization to models with overlapping generations of consumers will come later on. Moreover, we will, whenever we use the competitive equilibrium paradigm, assume that there is a "representative consumer". That is to say we think of it that there are a large (truly infinite, perhaps) number of consumers in the economy who are all identical. Prices of commodities will then have to adjust so that markets clear; this will typically mean (under appropriate strict concavity assumptions) that prices will make all these consumers make the same decisions: prices will have to adjust so that consumers do not interact. For example, the dynamic model without production gives a trivial allocation outcome: the consumer consumes the endowment of every product. The competitive mechanism ensures that this outcome is achieved by prices being set so that the consumer, when viewing prices as beyond his control, chooses to consume no more and no less than his endowments.

For a brief introduction, imagine that the production factors (capital and labor) were owned by many individual *households*, and that the technology to transform those factors into consumption goods was operated by *firms*. Then households' decisions would consist of the amount of factors to provide to firms, and the amount of consumption goods to purchase from them, while firms would have to choose their production volume and factor demand.

The device by which sellers and buyers (of factors and of consumption goods) are driven together is the *market*, which clearly brings with it the associated concept of *prices*. By equilibrium we mean a situation such that for some given *prices*, individual households' and firms' decisions show an aggregate consistency, i.e. the amount of factors that suppliers are willing to supply equals the amount that producers are willing to take, and the same for consumption goods - we say that *markets clear*. The word "competitive" indicates that we are looking at the perfect competition paradigm, as opposed to economies in which firms might have some sort of "market power".

Somewhat more formally, a competitive equilibrium is a vector of prices and quantities that satisfy certain properties related to the aggregate consistency of individual decisions mentioned above. These properties are:

1. Households choose quantities so as to maximize the level of utility attained given their "wealth" (factor ownership evaluated at the given prices). When making decisions, households take prices as given parameters. The maximum monetary value of goods that households are able to purchase given their wealth is called the budget constraint.

2. The quantity choice is "feasible". By this we mean that the aggregate amount of commodities that individual decision makers have chosen to demand can be produced with the available technology using the amount of factors that suppliers are willing to supply. Notice that this supply is in turn determined by the remuneration to factors, i.e. their price. Therefore this second condition is nothing but the requirement that *markets clear*.

3. Firms chose the production volume that maximizes their profits at the given prices.

For dynamic economic setups, we need to specify how trade takes place over time: are the economic agents using assets (and, if so, what kinds of assets)? Often, it will be possible to think of several different economic arrangements for the same physical environment that all give rise to the same final allocations. It will be illustrative to consider, for example, both the case when firms rent their inputs from consumers every period, and thus do not need an intertemporal perspective (and hence assets) to fulfill their profit maximization objective, and the case when they buy and own the long-lived capital they use in production, and hence need to consider the relative values of profits in different periods.

Also, in dynamic competitive equilibrium models, as in the maximization sections above, mathematically there are two alternative procedures: equilibria can be defined and analyzed in terms of (infinite) sequences, or they can be expressed recursively, using functions. We will look at both, starting with the former. For each approach, we will consider different specific arrangements, and we will proceed using examples: we will typically consider an example without production ("endowment economy") and the neoclassical growth model. Later applied chapters will feature many examples of other setups.

5.1 Sequential competitive equilibrium

The central question is the one of determining the set of commodities that are traded. The most straightforward extension of standard competitive analysis to dynamic models is perhaps the conceptually most abstract one: simply let goods be dated (so that, for example, in a one-good per date context, there is an infinite sequence of commodities: consumption at $t = 0$, consumption at $t = 1$, etc.) and, like in a static model, let the trade in all these commodities take place once and for all. We will call this setup the date-0 (or Arrow-Debreu-McKenzie) arrangement. In this arrangement, there is no need for assets. If, for example, a consumer needs to consume both in periods 0 and in future periods, the consumer would buy (rights to) future consumption goods at the beginning of time, perhaps in exchange for current labor services, or promises of future labor services. Any activity in the future would then be a mechanical carrying out of all the promises made at time zero.

An alternative setup is one with assets: we will refer to this case as one with sequential trade. In such a case, assets are used by one or more agents, and assets are traded every period. In such a case, there are nontrivial decisions made in every future period, unlike in the model with date-0 trade.

We will now, in turn, consider a series of example economies and, for each one, define equilibrium in a detailed way.

5.1.1 An endowment economy with date-0 trade

Let the economy have only one consumer with infinite life. There is no production, but the consumer is endowed with $\omega_t \in \Re$ units of the single consumption good at each date t. Notice that the absence of a production technology implies that the consumer is unable to move consumption goods across time; he must consume all his endowment in each period, or dispose of any balance. An economy without a production technology is called an *exchange economy*, since the only economic activity (besides consumption) that agents can undertake is trading. Let the consumer's utility from any given consumption path $\{c_t\}_{t=0}^{\infty}$ be given by

$$\sum_{t=0}^{\infty} \beta^t u\left(c_t\right).$$

The allocation problem in this economy is trivial. But imagine that we deceived the consumer into making him believe that he could actually engage in transactions to buy and sell consumption goods. Then, since in truth there is no other agent who could act as his counterpart, *market clearing* would require that prices are such that the consumer is willing to have exactly ω_t at every t.

We can see that this requires a specific price for consumption goods at each different point in time, i.e. the *commodities* here are consumption goods at different dates, and each commodity has its own price p_t. We can normalize ($p_0 = 1$) so that the prices will be relative to $t = 0$ consumption goods: a consumption good at t will cost p_t units of consumption goods at $t = 0$.

Given these prices, the value of the consumer's endowment is given by

$$\sum_{t=0}^{\infty} p_t \omega_t.$$

The value of his expenditures is

$$\sum_{t=0}^{\infty} p_t c_t$$

and the budget constraint requires that

$$\sum_{t=0}^{\infty} p_t c_t \leq \sum_{t=0}^{\infty} p_t \omega_t.$$

Notice that this assumes that trading in all *commodities* takes place at the same time: purchases and sales of consumption goods for every period are carried out at $t = 0$. This market structure is called an Arrow-Debreu-McKenzie, or date-0, market, as opposed to a sequential market structure, in which trading for each period's consumption good is undertaken in the corresponding period. Therefore in this example, we have the following:

Definition 5.1 *A* ***competitive equilibrium*** *is a vector of prices* $(p_t)_{t=0}^{\infty}$ *and a vector of quantities* $(c_t^*)_{t=0}^{\infty}$ *such that:*

1. $$(c_t^*)_{t=0}^{\infty} = \arg\max_{(c_t)_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t u(c_t) \right\}$$
$$s.t. \sum_{t=0}^{\infty} p_t c_t \leq \sum_{t=0}^{\infty} p_t \omega_t$$
$$c_t \geq 0 \ \forall t.$$

2. $c_t^* = \omega_t \ \forall t$ *(market clearing constraint).*

Notice, as mentioned earlier, that in this trivial case market clearing (condition 2) requires that the agent consumes exactly his endowment in each period, and this determines equilibrium prices.

Quantities are trivially determined here but prices are not. To find the price sequence that supports the quantities as a competitive equilibrium, simply use the first-order conditions from the consumer's problem. These are

$$\beta^t u'(\omega_t) = \lambda p_t \ \forall t,$$

where we have used the fact that equilibrium consumption c_t equals ω_t, and where λ denotes the Lagrange multiplier for the budget constraint. The multiplier can be eliminated to solve for any relative price, such as

$$\frac{p_t}{p_{t+1}} = \frac{1}{\beta} \frac{u'(\omega_t)}{u'(\omega_{t+1})}.$$

This equation states that the relative price of today's consumption in terms of tomorrow's consumption - the definition of the (gross) real interest rate - has to equal the marginal rate of substitution between these two goods, which in this case is inversely proportional to the discount rate and to the ratio of period marginal utilities. This price is expressed in terms of primitives and with it we have a complete solution for the competitive equilibrium for this economy (remember our normalization: $p_0 = 1$).

5.1.2 The same endowment economy with sequential trade

Let us look at the same *exchange economy*, but with a *sequential markets* structure. We allow 1-period loans, which carry an interest rate of

$$\underbrace{R_t}_{\text{gross rate}} \equiv 1 + \underbrace{r_t}_{\text{net rate}}$$

on a loan between periods $t-1$ and t. Let a_t denote the *net asset position* of the agent at time t, i.e. the net amount saved (lent) from last period.

Now we are allowing the agent to transfer wealth from one period to the next by lending 1-period loans to other agents. However, this is just a fiction as before, in the sense that since there is only one agent in the economy, there cannot actually be any loans outstanding (since lending requires both a lender and a borrower). Therefore the asset market will only clear if $a_t^* = 0 \ \forall t$, i.e. if the planned net asset holding is zero for every period.

With the new market structure, the agent faces not a single, but a sequence of budget constraints. His budget constraint in period t is given by:

$$\underbrace{c_t + a_{t+1}}_{\text{uses of funds}} = \underbrace{a_t R_t^* + \omega_t}_{\text{sources of funds}},$$

where R_t^* denotes the equilibrium interest rate that the agent takes as given. With this in hand, we have the following:

Definition 5.2 *A **competitive equilibrium** is a set of sequences* $\{c_t^*\}_{t=0}^{\infty}$, $\left\{a_{t+1}^*\right\}_{t=0}^{\infty}$, $\{R_t^*\}_{t=0}^{\infty}$ *such that:*

1. $$\left\{c_t^*, a_{t+1}^*\right\}_{t=0}^{\infty} = \underset{\{c_t,\, a_{t+1}\}_{t=0}^{\infty}}{\arg\max} \left\{\sum_{t=0}^{\infty} \beta^t u(c_t)\right\}$$
$$\text{s.t. } c_t + a_{t+1} = a_t R_t^* + \omega_t \ \forall t$$
$$c_t \geq 0 \ \forall t;\ a_0 = 0$$
$$\lim_{t\to\infty} a_{t+1} \left(\prod_{t=0}^{\infty} R_{t+1}\right)^{-1} = 0 \text{ (no-Ponzi-game condition).}$$

2. *Feasibility constraint:* $a_t^* = 0 \ \forall t$ *(asset market clearing).*

3. $c_t^* = \omega_t \ \forall t$ *(goods market clearing).*

Notice that the third condition necessarily follows from the first and second ones, by Walras's law: if $n-1$ markets clear in each period, then the n^{th} one will clear as well.

To determine quantities is as trivial here (with the same result) as in the date-0 world. Prices, i.e. interest rates, are again available from the first-order condition for saving, the consumer's Euler equation, evaluated at $c_t^* = \omega_t$:

$$u'(\omega_t) = \beta u'(\omega_{t+1}) R_{t+1}^*,$$

so that

$$R_{t+1}^* = \frac{1}{\beta} \frac{u'(\omega_t)}{u'(\omega_{t+1})}.$$

Not surprisingly, this expression coincides with the real interest rate in the date-0 economy.

5.1.3 The neoclassical growth model with date-0 trade

Next we will look at an application of the definition of competitive equilibrium to the neoclassical growth model. We will first look at the definition of competitive equilibrium with a date-0 market structure, and then at the sequential markets structure.

The assumptions in our version of the neoclassical growth model are as follows:

1. The consumer is endowed with 1 unit of "time" each period, which he can allocate between labor and leisure.

2. The utility derived from the consumption and leisure stream $\{c_t,\, 1-n_t\}_{t=0}^{\infty}$ is given by
$$U\left(\{c_t,\, 1-n_t\}_{t=0}^{\infty}\right) = \sum_{t=0}^{\infty} \beta^t u\left(c_t\right).$$
That is, we assume for the moment that leisure is not valued; equivalently, labor supply bears no utility cost. We also assume that $u\left(\cdot\right)$ is strictly increasing and strictly concave.

3. The consumer owns the capital, which he rents to firms in exchange for r_t units of the consumption good at t per unit of capital rented. Capital depreciates at rate δ each period.

4. The consumer rents his labor services at t to the firm for a unit rental (or wage) rate of w_t.

5. The production function of the consumption/investment good is $F(K, n)$; F is strictly increasing in each argument, concave, and homogeneous of degree 1.

The following are the prices involved in this market structure:

- Price of consumption good at every t: p_t

  p_t: intertemporal relative prices; if $p_0 = 1$, then p_t is the price of consumption goods at t relative to (in terms of) consumption goods at $t = 0$.

- Price of capital services at t: $p_t r_t$

  r_t: rental rate; price of capital services at t relative to (in terms of) consumption goods at t.

- Price of labor: $p_t w_t$

  w_t: wage rate; price of labor at t relative to (in terms of) consumption goods at t.

Definition 5.3 *A* ***competitive equilibrium*** *is a set of sequences:*

Prices: $\{p_t^*\}_{t=0}^{\infty}$, $\{r_t^*\}_{t=0}^{\infty}$, $\{w_t^*\}_{t=0}^{\infty}$

Quantities: $\{c_t^*\}_{t=0}^{\infty}$, $\left\{K_{t+1}^*\right\}_{t=0}^{\infty}$, $\{n_t^*\}_{t=0}^{\infty}$ *such that*

1. $\{c_t^*\}_{t=0}^{\infty}$, $\left\{K_{t+1}^*\right\}_{t=0}^{\infty}$, $\{n_t^*\}_{t=0}^{\infty}$ *solve the consumer's problem:*

$$\left\{c_t^*,\, K_{t+1}^*,\, n_t^*\right\}_{t=0}^{\infty} = \underset{\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}}{\arg\max} \left\{\sum_{t=0}^{\infty} \beta^t u\left(c_t\right)\right\}$$

$$s.t.\ \sum_{t=0}^{\infty} p_t^*\left[c_t + K_{t+1}\right] = \sum_{t=0}^{\infty} p_t^*\left[r_t^* K_t + (1-\delta) K_t + n_t w_t^*\right]$$

$$c_t \geq 0\ \forall t, \quad k_0 \text{ given.}$$

At every period t, capital is quoted in the same price as the consumption good. As for labor, recall that we have assumed that it has no utility cost. Therefore $w_t > 0$ will imply that the consumer supplies all his time endowment to the labor market: $w_t > 0 \Rightarrow n_t^* = 1\ \forall t$.

2. $\{K_t^*\}_{t=0}^{\infty}$, $\{n_t^*\}_{t=0}^{\infty}$ *solve the firms' problem:*

$$\forall t : (K_t^*, 1) = \underset{K_t,\, n_t}{\arg\max} \left\{p_t^* F\left(K_t, n_t\right) - p_t^* r_t^* K_t - p_t^* w_t^* n_t\right\}$$

The firm's decision problem involves just a one-period choice - it is not of a dynamical nature (for example, we could imagine that firms live for just one period). All of the model's dynamics come from the consumer's capital accumulation problem.

This condition may equivalently be expressed as follows: $\forall t : (r_t^*, w_t^*)$ satisfy:

$$\begin{aligned} r_t^* &= F_K\left(K_t^*, 1\right) \qquad (5.1) \\ w_t^* &= F_n\left(K_t^*, 1\right). \end{aligned}$$

Notice that this shows that if the production function $F(K, n)$ is increasing in n, then $n_t^* = 1$ follows.

3. Feasibility (market clearing):

$$c_t^* + K_{t+1}^* = F\left(K_t^*, 1\right) + (1-\delta)\, K_t^*.$$

This is known as the one-sector neoclassical growth model, since only one type of goods is produced, that can be used either for consumption in the current period or as capital in the following. There is also a vast literature on multi-sector neoclassical growth models, in which each type of physical good is produced with a different production technology, and capital accumulation is specific to each technology.

Let us now characterize the equilibrium. We first study the consumer's problem by deriving his intertemporal first-order conditions. Differentiating with respect to c_t, we obtain

$$c_t : \beta^t u'\left(c_t^*\right) = p_t^* \lambda^*,$$

where λ^* is the Lagrange multiplier corresponding to the budget constraint. Since the market structure that we have assumed consists of date-0 markets, there is only one budget and hence a unique multiplier.

Consumption at $t+1$ obeys

$$c_{t+1} : \beta^{t+1} u'\left(c_{t+1}^*\right) = p_{t+1}^* \lambda^*.$$

Combining the two we arrive at

$$\frac{p_t^*}{p_{t+1}^*} = \frac{1}{\beta} \frac{u'\left(c_t^*\right)}{u'\left(c_{t+1}^*\right)}. \tag{5.2}$$

We can, as before, interpret $\frac{p_t^*}{p_{t+1}^*}$ as the real interest rate, and $\frac{1}{\beta}\frac{u'\left(c_t^*\right)}{u'\left(c_{t+1}^*\right)}$ as the marginal rate of substitution of consumption goods between t and $t+1$.

Differentiating with respect to capital, one sees that

$$K_{t+1} : \lambda^* p_t^* = \lambda^* p_{t+1}^* \left[r_{t+1}^* + (1-\delta)\right].$$

Therefore,

$$\frac{p_t^*}{p_{t+1}^*} = r_{t+1}^* + 1 - \delta.$$

Using condition (5.1), we also find that

$$\frac{p_t^*}{p_{t+1}^*} = F_K\left(K_{t+1}^*, 1\right) + 1 - \delta. \tag{5.3}$$

The expression $F_K\left(K_{t+1}^*, 1\right) + (1-\delta)$ is the marginal return on capital: the marginal rate of technical substitution (transformation) between c_t and c_{t+1}. Combining expressions (5.2) and (5.3), we see that

$$u'(c_t^*) = \beta u'(c_{t+1}^*)\left[F_K\left(K_{t+1}^*, 1\right) + 1 - \delta\right]. \tag{5.4}$$

Notice now that (5.4) is nothing but the Euler Equation from the planner's problem. Therefore a competitive equilibrium allocation satisfies the optimality conditions for the centralized economy: the competitive equilibrium is optimal. You may recognize this as the First Welfare Theorem. We have assumed that there is a single consumer, so in this case Pareto-optimality just means utility maximization. In addition, as we will see later, with the appropriate assumptions on $F(K, n)$ (namely, non-increasing returns to scale), an optimum can be supported as a competitive equilibrium, which is the result of the Second Welfare Theorem.

5.1.4 The neoclassical growth model with sequential trade

The following are the prices involved in this market structure:

- Price of capital services at t: R_t

  R_t: rental rate; price of capital services at t relative to (in terms of) consumption goods at t.

  Just for the sake of variety, we will now assume that R_t is the return on capital net of the depreciation costs. That is, with the notation used before, $R_t \equiv r_t + 1 - \delta$.

- Price of labor: w_t

  w_t: wage rate; price of labor at t relative to (in terms of) consumption goods at t.

Definition 5.4 *A* ***competitive equilibrium*** *is a sequence* $\left\{R_t^*,\, w_t^*,\, c_t^*,\, K_{t+1}^*,\, n_t^*\right\}_{t=0}^{\infty}$ *such that:*

1. $\left\{c_t^*,\, K_{t+1}^*,\, n_t^*\right\}_{t=0}^{\infty}$ *solves the consumer's problem:*

$$\left\{c_t^*,\, K_{t+1}^*,\, n_t^*\right\}_{t=0}^{\infty} = \underset{\left\{c_t,\, K_{t+1},\, n_t\right\}_{t=0}^{\infty}}{\arg\max} \left\{\sum_{t=0}^{\infty} \beta^t u(c_t)\right\}$$

$$\text{s.t. } c_t + K_{t+1} = K_t R_t^* + n_t w_t^*$$

$$k_0 \text{ given and a no-Ponzi-game condition.}$$

(Note that accumulating K_{t+1} *is analogous to lending at* t*.)*

2. $\left\{K_{t+1}^*,\, n_t^*\right\}_{t=0}^{\infty}$ *solves the firms' problem:*

$$\forall t : (K_t^*, 1) = \underset{K_t,\, n_t}{\arg\max} \left\{F\left(K_t, n_t\right) - R_t^* K_t + (1-\delta) K_t - w_t^* n_t\right\}.$$

3. *Market clearing (feasibility):*

$$\forall t : c_t^* + K_{t+1}^* = F\left(K_t^*, 1\right) + (1-\delta) K_t^*.$$

The way that the rental rate has been presented now can be interpreted as saying that the firm manages the capital stock, funded by loans provided by the consumers. However, the capital accumulation decisions are still in the hands of the consumer (this might also be modeled in a different way, as we shall see later).

Let us solve for the equilibrium elements. As before, we start with the consumer's problem:

$$c_t : \beta^t u'(c_t^*) = \beta^t \lambda_t^*.$$

With the current market structure, the consumer faces a sequence of budget constraints, and hence a sequence of Lagrange multipliers $\{\lambda_t^*\}_{t=0}^{\infty}$. We also have

$$c_{t+1} : \beta^{t+1} u'\left(c_{t+1}^*\right) = \beta^{t+1} \lambda_{t+1}^*.$$

Then

$$\frac{\lambda_t^*}{\lambda_{t+1}^*} = \frac{u'(c_t^*)}{u'\left(c_{t+1}^*\right)}. \tag{5.5}$$

Differentiation with respect to capital yields

$$K_{t+1} : \beta^t \lambda_t^* = \beta^{t+1} R_{t+1}^* \lambda_{t+1}^*,$$

so that

$$\frac{\lambda_t^*}{\lambda_{t+1}^*} = \beta R_{t+1}^*. \tag{5.6}$$

Combining expressions (5.5) and (5.6), we obtain

$$\frac{u'(c_t^*)}{u'\left(c_{t+1}^*\right)} = \beta R_{t+1}^*. \tag{5.7}$$

From Condition 2 of the definition of competitive equilibrium,

$$R_t^* = F_k(K_t^*, 1) + 1 - \delta. \tag{5.8}$$

Therefore, combining (5.7) and (5.8) we obtain:

$$u'(c_t^*) = \beta u'\left(c_{t+1}^*\right)\left[F_k(K_t^*, 1) + 1 - \delta\right].$$

This, again, is identical to the planner's Euler equation. It shows that the sequential market equilibrium is the same as the Arrow-Debreu-McKenzie date-0 equilibrium and both are Pareto-optimal.

5.2 Recursive competitive equilibrium

Recursive competitive equilibrium uses the recursive concept of treating all maximization problems as split into decisions concerning today versus the entire future. As such, this concept thus has no room for the idea of date-0 trading: it requires sequential trading.

Instead of having sequences (or vectors), a recursive competitive equilibrium is a set of functions - quantities, utility levels, and prices, as functions of the "state": the relevant initial condition. As in dynamic programming, these functions allow us to say what will happen in the economy for every specific consumer, given an arbitrary choice of the initial state.

As above, we will state the definitions and then discuss their ramifications in the context of a series of examples, beginning with a treatment of the neoclassical growth model.

5.2.1 The neoclassical growth model

Let us assume again that the time endowment is equal to 1, and that leisure is not valued. Recall the central planner's problem that we analyzed before:

$$\begin{aligned} V(K) &= \max_{c,\, K' \geq 0} \{u(c) + \beta V(K')\} \\ \text{s.t.} \;\; & c + K' = F\left(K,\, 1\right) + \left(1 - \delta\right) K. \end{aligned}$$

In the decentralized recursive economy, the individual's budget constraint will no longer be expressed in terms of physical units, but in terms of sources and uses of funds at the going market prices. In the sequential formulation of the decentralized problem, these take the form of sequences of factor remunerations: $\{R_t,\, w_t\}_{t=0}^{\infty}$, with the equilibrium levels given by

$$\begin{aligned} R_t^* &= F_K\left(K_t^*,\, 1\right) + 1 - \delta \\ w_t^* &= F_n\left(K_t^*,\, 1\right). \end{aligned}$$

Notice that both are a function of the (aggregate) level of capital (with aggregate labor supply normalized to 1). In dynamic programming terminology, what we have is a law of motion for factor remunerations as a function of the aggregate level of capital in the economy. If $\bar{K}$ denotes the (current) aggregate capital stock, then

$$\begin{aligned} R &= R(\bar{K}) \\ w &= w(\bar{K}). \end{aligned}$$

Therefore, the budget constraint in the decentralized dynamic programming problem reads

$$c + K' = R(\bar{K})K + w(\bar{K}). \tag{5.9}$$

The previous point implies that when making decisions, two variables are key to the agent: his own level of capital, K, and the aggregate level of capital, $\bar{K}$, which will determine his income. So the correct "syntax" for writing down the dynamic programming problem is:

$$V\left(K, \bar{K}\right) = \max_{c,\, K' \geq 0} \left\{u(c) + \beta V\left(K', \bar{K}'\right)\right\}, \tag{5.10}$$

where the state variables for the consumer are K and $\bar{K}$).

We already have the objective function that needs to be maximized and one of the restrictions, namely the budget constraint. Only $\bar{K}'$ is left to be specified. The economic interpretation of this is that we must determine the agent's *perceived* law of motion of aggregate capital. We assume that he will perceive this law of motion as a function of the aggregate level of capital. Furthermore, his perception will be *rational* - it will correctly correspond to the actual law of motion:

$$\bar{K}' = G(\bar{K}), \tag{5.11}$$

where G is a result of the economy's, that is, the representative agent's equilibrium capital accumulation decisions.

Putting (5.9), (5.10) and (5.11) together, we write down the consumer's complete dynamic problem in the decentralized economy:

$$V\left(K, \bar{K}\right) = \max_{c,\, K' \geq 0} \left\{u(c) + \beta V\left(K', \bar{K}'\right)\right\} \tag{5.12}$$

$$\text{s.t.} \quad c + K' = R(\bar{K})K + w(\bar{K})$$
$$\bar{K}' = G(\bar{K}).$$

(5.12) is the recursive competitive equilibrium functional equation. The solution will yield a policy function for the individual's law of motion for capital:

$$K' = g(K, \bar{K}) = \underset{K' \in [0, R(\bar{K})K + w(\bar{K})]}{\arg\max} \left\{u[R(\bar{K})K + w(\bar{K}) - \bar{K}'] + \beta V\left(K', \bar{K}'\right)\right\}$$

$$\text{s.t.} \quad \bar{K}' = G(\bar{K}).$$

We can now address the object of our study:

Definition 5.5 *A* ***recursive competitive equilibrium*** *is a set of functions:*

Quantities: $G(\bar{K})$, $g(K, \bar{K})$

Lifetime utility level: $V\left(K, \bar{K}\right)$

Prices: $R(\bar{K})$, $w(\bar{K})$ *such that*

1. $V\left(K, \bar{K}\right)$ *solves (5.12) and* $g(K, \bar{K})$ *is the associated policy function.*

2. *Prices are competitively determined:*

$$\begin{aligned} R(\bar{K}) &= F_K(\bar{K}, 1) + 1 - \delta \\ w(\bar{K}) &= F_n(\bar{K}, 1). \end{aligned}$$

In the recursive formulation, prices are stationary functions, rather than sequences.

3. *Consistency is satisfied:*

$$G(\bar{K}) = g(\bar{K}, \bar{K}) \ \forall \bar{K}.$$

The third condition is the distinctive feature of the recursive formulation of competitive equilibrium. The requirement is that, whenever the individual consumer is endowed with a level of capital equal to the aggregate level (for example, only one single agent in the economy owns all the capital, or there is a measure one of agents), his own individual behavior will exactly mimic the aggregate behavior. The term *consistency* points out the fact that the aggregate law of motion *perceived* by the agent must be *consistent* with the actual behavior of individuals. Consistency in the recursive framework corresponds to the idea in the sequential framework that consumers' chosen sequences of, say, capital, have to satisfy their first-order conditions given prices that are determined from firms' first-order conditions *evaluated using the same sequences of capital.*

None of the three conditions defining a recursive competitive equilibrium mentions market clearing. Will markets clear? That is, will the following equality hold?

$$\bar{c} + \bar{K}' = F(\bar{K},\, 1) + (1 - \delta)\, \bar{K},$$

where $\bar{c}$ denotes aggregate consumption. To answer this question, we may make use of the Euler Theorem. If the production technology exhibits constant returns to scale (that is, if the production function is homogeneous of degree 1), then that theorem delivers:

$$F(\bar{K},\, 1) + (1 - \delta)\, \bar{K} = R(\bar{K})\bar{K} + w(\bar{K}).$$

In economic terms, there are zero profits: the product gets exhausted in factor payment. This equation, together with the consumer's budget constraint evaluated in equilibrium ($K = \bar{K}$) implies market clearing.

Completely solving for a recursive competitive equilibrium involves more work than solving for a sequential equilibrium, since it involves solving for the functions V and g, which specify "off-equilibrium" behavior: what the agent would do if he were different from the representative agent. This calculation is important in the sense that in order to justify the equilibrium behavior we need to see that the postulated, chosen path, is not worse than any other path. $V(K, \bar{K})$ precisely allows you to evaluate the future consequences for these behavioral alternatives, thought of as one-period deviations. Implicitly this is done with the sequential approach also, although in that approach one typically simply derives the first-order (Euler) equation and imposes $K = \bar{K}$ there. Knowing that the F.O.C. is sufficient, one does not need to look explicitly at alternatives.

The known parametric cases of recursive competitive equilibria that can be solved fully include the following ones: (i) logarithmic utility (additive logarithms of consumption and leisure, if leisure is valued), Cobb-Douglas production, and 100% depreciation; (ii) isoelastic utility and linear production; and (iii) quadratic utility and linear production. It is also possible to show that, when utility is isoelastic (and no matter what form the production function takes), one obtains decision rules of the form $g(K, \bar{K}) = \lambda(\bar{K})K + \mu(\bar{K})$, where the two functions λ and μ satisfy a pair of functional equations whose solution depends on the technology and on the preference parameters. That is, the individual decision rules are linear in K, the agent's own holdings of capital.

More in the spirit of solving for sequential equilibria, one can solve for recursive competitive equilibrium less than fully by ignoring V and g and only solve for G, using the competitive equilibrium version of the functional Euler equation. It is straightforward to show, using the envelope theorem as above in the section on dynamic programming, that this functional equation reads

$$\begin{aligned} u'\left(R(\bar{K})K + w(\bar{K}) - g(K,\bar{K})\right) &= \beta u'\left(R(G(\bar{K}))g(K,\bar{K}) + w(G(\bar{K})) - \right. \\ &\quad \left. - g(g(K,\bar{K}), G(\bar{K}))\right)\left(F_1(G(K),1) + 1 - \delta\right) \ \forall K, \bar{K}. \end{aligned}$$

Using the Euler Theorem and consistency ($K = \bar{K}$) we now see that this functional equation becomes

$$\begin{aligned} u'\left(F(\bar{K},1) + (1-\delta)\bar{K} - G(\bar{K})\right) &= \beta u'\left(F(G(\bar{K}),1) + (1-\delta)G(\bar{K}) - \right. \\ &\quad \left. - G(G(\bar{K}))\right)\left(F_1(G(\bar{K}),1) + 1 - \delta\right) \ \forall \bar{K}, \end{aligned}$$

which corresponds exactly to the functional Euler equation in the planning problem. We have thus shown that the recursive competitive equilibrium produces optimal behavior.

5.2.2 The endowment economy with one agent

Let the endowment process be stationary: $\omega_t = \omega$, $\forall t$. The agent is allowed to save in the form of loans (or assets). His net asset position at the beginning of the period is given by a. Asset positions need to cancel out in the aggregate: $\overline{a} = 0$, since for every lender there must be a borrower. The definition of a recursive equilibrium is now as follows.

Definition 5.6 *A* ***recursive competitive equilibrium*** *is a set of functions $V(a)$, $g(a)$, R such that*

1. *$V(a)$ solves the consumer's functional equation:*

$$\begin{aligned} V(a) &= \max_{c \geq 0,\, a'} \{u(c) + \beta V(a')\} \\ s.t.\ c + a' &= aR + \omega. \end{aligned}$$

2. *Consistency:*

$$g(0) = 0.$$

The consistency condition in this case takes the form of requiring that the agent that has a null initial asset position keep this null balance. Clearly, since there is a unique agent then asset market clearing requires $a = 0$. This condition will determine R as the return on assets needed to sustain this equilibrium. Notice also that R is not really a function - it is a constant, since the aggregate net asset position is zero.

Using the functional Euler equation, which of course can be derived here as well, it is straightforward to see that R has to satisfy

$$R = \frac{1}{\beta},$$

since the u' terms cancel. This value induces agents to save zero, if they start with zero assets. Obviously, the result is the same as derived using the sequential equilibrium definition.

5.2.3 An endowment economy with two agents

Assume that the economy is composed of two agents who live forever. Agent i derives utility from a given consumption stream $\{c_t^i\}_{t=0}^{\infty}$ as given in the following formula:

$$U_i\left(\{c_t^i\}_{t=0}^{\infty}\right) = \sum_{t=0}^{\infty} \beta_i^t u_i\left(c_t^i\right), \; i = 1, 2.$$

Endowments are stationary:

$$\omega_t^i = \omega^i \; \forall t, \; i = 1, 2.$$

Total resource use in this economy must obey:

$$c_t^1 + c_t^2 = \omega^1 + \omega^2 \; \forall t.$$

Clearing of the asset market requires that:

$$\overline{a}_t \equiv a_t^1 + a_t^2 = 0 \; \forall t.$$

Notice this implies $a_t^1 = -a_t^2$; that is, at any point in time it suffices to know the asset position of one of the agents to know the asset position of the other one as well. Denote $A_1 \equiv a^1$. This is the relevant aggregate state variable in this economy (the time subscript is dropped to adjust to dynamic programming notation). Claiming that it is a state variable amounts to saying that the distribution of asset holdings will matter for prices. This claim is true except in special cases (as we shall see below), because whenever marginal propensities to save out of wealth are not the same across the two agents (either because they have different utility functions or because their common utility function makes the propensity depend on the wealth level), different prices are required to make total savings be zero, as equilibrium requires.

Finally, let q denote the current price of a one-period bond: $q_t = \frac{1}{R_{t,\,t+1}}$. Also, in what follows, subscript denotes the type of agent. We are now ready to state the following:

Definition 5.7 *A* ***recursive competitive equilibrium*** *of the two-agent economy is a set of functions:*

Quantities: $g_1\left(a_1,\, A_1\right),\, g_2\left(a_2,\, A_1\right),\, G\left(A_1\right)$

Lifetime utility levels: $V_1(a_1, A_1), V_2(a_2, A_1)$

Prices: $q(A_1)$ *such that*

1. $V_i(a_i, A_1)$ *is the solution to consumer i's problem:*

$$\begin{aligned} V_i(a_i, A_1) &= \max_{c^i \geq 0, a'_i} \left\{ u_i\left(c^i\right) + \beta_i V_i(a'_i, A'_1) \right\} \\ s.t.\ c^i + a'_i q(A_1) &= a_i + \omega_i. \\ A'_1 &= G(A_1) \ \to \textit{perceived law of motion for } A_1. \end{aligned}$$

   The solution to this functional equation delivers the policy function $g_i(a_i, A_1)$.

2. *Consistency:*

$$\begin{aligned} G(A_1) &= g_1(A_1, A_1) \quad \forall A_1 \\ -G(A_1) &= g_2(-A_1, A_1) \ \forall A_1. \end{aligned}$$

The second condition implies asset market clearing:

$$g_1(A_1, A_1) + g_2(-A_1, A_1) = G(A_1) - G(A_1) = 0.$$

Also note that q is the variable that will adjust for consistency to hold.

For this economy, it is not as easy to find analytical solutions, except for special parametric assumptions. We will turn to those now. We will, in particular, consider the following question: under what conditions will q be constant (that is, independent of the wealth distribution characterized by A_1)?

The answer is that, as long as $\beta_1 = \beta_2$ and u_i is strictly concave, q will equal β and thus not depend on A_1. This is easily shown by guessing and verifying; the remainder of the functions are as follows: $g_i(a, A_1) = a$ for all i and (a, A_1) and $G(A_1) = A_1$ for all A_1.

5.2.4 Neoclassical production again, with capital accumulation by firms

Unlike in the previous examples (recall the discussion of competitive equilibrium with the sequential and recursive formulations), we will now assume that firms are the ones that make capital accumulation decisions in this economy. The (single) consumer owns stock in the firms. In addition, instead of labor, we will have "land" as the second factor of production. Land will be owned by the firm.

The functions involved in this model are the dynamic programs of both the consumer and the firm:

$\bar{K}' = G(\bar{K})$ — aggregate law of motion for capital.

$q(\bar{K})$ — current price of next period's consumption $\left(\frac{1}{\text{return on stocks}}\right)$.

$V_c\left(a,\, \bar{K}\right)$ — consumer's indirect utility as function of $\bar{K}$ and his wealth a.

$a' = g_c\left(a,\, \bar{K}\right)$ — policy rule associated with $V_c\left(a,\, \bar{K}\right)$.

$V_f\left(K,\, \bar{K}\right)$ — market value (in consumption goods), of a firm with K units of initial capital, when the aggregate stock of capital is $\bar{K}$.

$K' = g_f\left(K,\, \bar{K}\right)$ — policy rule associated with $V_f\left(K,\, \bar{K}\right)$.

The dynamic programs of the different agents are as follows:

1. The consumer:

$$V_c\left(a,\, \bar{K}\right) = \max_{c \geq 0,\, a'} \left\{u(c) + \beta V_c\left(a',\, \bar{K}'\right)\right\} \tag{5.13}$$

$$\text{s.t. } c + q\left(\bar{K}\right) a' = a$$

$$\bar{K}' = G\left(\bar{K}\right).$$

The solution to this dynamic program produces the policy rule

$$a' = g_c\left(a,\, \bar{K}\right).$$

2. The firm:

$$V_f\left(K,\, \bar{K}\right) = \max_{K'} \left\{F\left(K,\, 1\right) + \left(1 - \delta\right) K - K' + q\left(\bar{K}\right) V_f\left(K',\, \bar{K}'\right)\right\} \tag{5.14}$$

$$\text{s.t. } \bar{K}' = G\left(\bar{K}\right).$$

The solution to this dynamic program produces the policy rule

$$K' = g_f\left(K,\, \bar{K}\right).$$

We are now ready for the equilibrium definition.

Definition 5.8 *A* ***recursive competitive equilibrium*** *is a set of functions*

Quantities: $g_c\left(a,\, \bar{K}\right),\, g_f\left(K,\, \bar{K}\right),\, G\left(\bar{K}\right)$

Lifetime utility levels, values: $V_c\left(a,\, \bar{K}\right),\, V_f\left(K,\, \bar{K}\right)$

Prices: $q\left(\bar{K}\right)$ *such that*

1. *$V_c\left(a,\,\bar{K}\right)$ and $g_c\left(a,\,\bar{K}\right)$ are the value and policy functions, respectively, solving (5.13).*

2. *$V_f\left(K,\,\bar{K}\right)$ and $g_f\left(K,\,\bar{K}\right)$ are the value and policy functions, respectively, solving (5.14).*

3. *Consistency 1: $g_f\left(\bar{K},\,\bar{K}\right) = G\left(\bar{K}\right)$ for all $\bar{K}$.*

4. *Consistency 2: $g_c\left[V_f\left(\bar{K},\,\bar{K}\right),\,\bar{K}\right] = V_f\left[G\left(\bar{K}\right),\,G\left(\bar{K}\right)\right]\ \forall\bar{K}$.*

The consistency conditions can be understood as follows. The last condition requires that the consumer ends up owning 100% of the firm next period whenever he started up owning 100% of it. Notice that if the consumer starts the period owning the whole firm, then the value of a (his wealth) is equal to the market value of the firm, given by $V_f(\cdot)$. That is,

$$a = V_f\left(K,\,\bar{K}\right). \tag{5.15}$$

The value of the firm next period is given by

$$V_f(K',\ \bar{K}').$$

To assess this value, we need K' and $\bar{K}'$. But these come from the respective laws of motion:

$$V_f(K',\bar{K}') = V_f\left[g_f(K,\ \bar{K}),\,G(\bar{K})\right].$$

Now, requiring that the consumer owns 100% of the firm in the next period amounts to requiring that his desired asset accumulation, a', coincide with the value of the firm next period:

$$a' = V_f\left[g_f(K,\ \bar{K}),\,G(\bar{K})\right].$$

But a' follows the policy rule $g_c(a,\,\bar{K})$. A substitution then yields

$$g_c(a,\bar{K}) = V_f\left[g_f(K,\ \bar{K}),\,G(\bar{K})\right]. \tag{5.16}$$

Using (5.15) to replace a in (5.16), we obtain

$$g_c\left[V_f\left(K,\,\bar{K}\right),\bar{K}\right] = V_f\left[g_f(K,\ \bar{K}),\,G(\bar{K})\right]. \tag{5.17}$$

The consistency condition is then imposed with $K = \bar{K}$ in (5.17) (and using the "Consistency 1" condition $g_f\left[\bar{K},\bar{K}\right] = G\left[\bar{K}\right]$), yielding

$$g_c\left[V_f\left(\bar{K},\,\bar{K}\right),\bar{K}\right] = V_f\left[G(\bar{K}),\,G(\bar{K})\right].$$

To show that the allocation resulting from this definition of equilibrium coincides with the allocation we have seen earlier (e.g., the planning allocation), one would have to derive functional Euler equations for both the consumer and the firm and simplify them. We leave it as an exercise to verify that the outcome is indeed the optimal one.

Chapter 6

Uncertainty

Our program of study will comprise the following three topics:

1. Examples of common stochastic processes in macroeconomics

2. Maximization under uncertainty

3. Competitive equilibrium under uncertainty

The first one is closely related to time series analysis. The second and the third one are a generalization of the tools we have already introduced to the case where the decision makers face uncertainty.

Before proceeding with this chapter, it may be advisable to review the basic notation and terminology associated with stochastic processes presented in the appendix.

6.1 Examples of common stochastic processes in macroeconomics

The two main types of modelling techniques that macroeconomists make use of are:

- Markov chains

- Linear stochastic difference equations

6.1.1 Markov chains

Definition 6.1 *Let $x_t \in X$, where $X = \overline{x}_1, \overline{x}_2, ..., \overline{x}_n$ is a finite set of values. A **stationary Markov chain** is a stochastic process $\{x_t\}_{t=0}^{\infty}$ defined by X, a transition matrix $\underset{n\times n}{P}$, and an initial probability distribution $\underset{1\times n}{\pi_0}$ for x_0 (the first element in the stochastic process).*

The elements of $\underset{n\times n}{P}$ represent the following probabilities:

$$P_{ij} = \Pr[x_{t+1} = \overline{x}_j | x_t = \overline{x}_i].$$

Notice that these probabilities are independent of time. We also have that the probability two periods ahead is given by

$$\begin{aligned} \Pr[x_{t+2} = \overline{x}_j | x_t = \overline{x}_i] &= \sum_{k=1}^{n} P_{ik} P_{kj} \\ &\equiv [P^2]_{i,j}, \end{aligned}$$

where $[P^2]_{i,j}$ denotes the $(i,\ j)^{\text{th}}$ entry of the matrix P^2.

Given π_0, π_1 is the probability distribution of $x_1$as of time $t = 0$ and it is given by

$$\pi_1 = \pi_0 P.$$

Analogously,

$$\begin{aligned} \pi_2 &= \pi_0 P^2 \\ \vdots &= \vdots \\ \pi_t &= \pi_0 P^t \end{aligned}$$

and also

$$\pi_{t+1} = \pi_t P.$$

Definition 6.2 *A* ***stationary*** *(or* ***invariant****) distribution for P is a probability vector π such that*

$$\pi = \pi P.$$

A stationary distribution then satisfies

$$\pi I = \pi P,$$

where I is identity matrix and

$$\begin{aligned} \pi - \pi P &= 0 \\ \pi[I - P] &= 0. \end{aligned}$$

That is, π is an eigenvector of P, associated with the eigenvalue $\lambda = 1$.

Example 6.3

(i) $P = \begin{pmatrix} .7 & .3 \\ .6 & .4 \end{pmatrix} \Rightarrow \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} .7 & .3 \\ .6 & .4 \end{pmatrix}$ *You should verify that* $\pi = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \end{pmatrix}$.

(ii) $P = \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \Rightarrow \pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

(iii) $P = \begin{pmatrix} 1 & 0 \\ .1 & .9 \end{pmatrix} \Rightarrow \pi = \begin{pmatrix} 1 & 0 \end{pmatrix}$. *The first state is said to be "absorbing".*

(iv) $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow \pi = \begin{pmatrix} a & 1-a \end{pmatrix},\ a \in [0,\, 1]$. *In this last case, there is a continuum of invariant distributions.*

The question now is whether π_t converges, in some sense, to a number π_∞ as $t \to \infty$, which would mean that $\pi_\infty = \pi_\infty P$ and if so, whether π_∞ depends on the initial condition π_0. If the answers to these two questions are "Yes" and "No", respectively, then the stochastic process is said to be "asymptotically stationary", with a unique invariant distribution. Fortunately, we can borrow the following result for sufficient conditions for asymptotic stationarity:

Theorem 6.4 *P has a unique invariant distribution (and is asymptotically stationary) if $P_{ij} > 0\ \forall i,\ \forall j$.*

6.1.2 Linear stochastic difference equations

Let $x_t \in \Re^n,\ \ w_t \in \Re^m$,

$$x_{t+1} = \underset{n\times n}{A}\, x_t + \underset{n\times n}{C}\, w_{t+1}.$$

We normally assume

$$\begin{aligned} E_t\left[w_{t+1}\right] &= E_t\left[w_{t+1}\left|w_t,\, w_{t-1},\, ...\right.\right] = 0 \\ E_t\left[w_{t+1}w'_{t+1}\right] &= I. \end{aligned}$$

Example 6.5 (AR(1) process) *Let*

$$y_{t+1} = \rho y_t + \varepsilon_{t+1} + b$$

and assume

$$\begin{aligned} E_t\left[\varepsilon_{t+1}\right] &= 0 \\ E_t\left[\varepsilon_{t+1}^2\right] &= \sigma^2 \\ E_t\left[\varepsilon_{t+k}\varepsilon_{t+k+1}\right] &= 0. \end{aligned}$$

Even if y_0 is known, the $\{y_t\}_{t=0}^{\infty}$ process will not be stationary in general. However, the process may become stationary as $t \to \infty$. By repeated substitution, we get

$$E_0\left[y_t\right] = \rho^t y_0 + \frac{b}{1-\rho}\left(1-\rho^t\right)$$

$$|\rho| < 1 \Rightarrow \lim_{t\to\infty} E_0[y_t] = \frac{b}{1-\rho}.$$

Then, the process will be stationary if $|\rho| < 1$*. Similarly, the autocovariance function is given by*

$$\gamma(t,\, k) \equiv E_0[(y_t - E[y_t])(y_{t-k} - E[y_{t-k}])] = \sigma^2 \rho^k \frac{1-\rho^{t-k}}{1-\rho^2}$$

$$|\rho| < 1 \Rightarrow \lim_{t\to\infty} \gamma(t,\, k) = \frac{\sigma^2}{1-\rho^2}\rho^k.$$

The process is asymptotically weakly stationary if $|\rho| < 1$*.*

We can also regard x_0 (or y_0, in the case of an AR(1) process) as drawn from a distribution with mean μ_0 and covariance $E\left[(x_0 - \mu_0)(x_0 - \mu_0)'\right] \equiv \Gamma_0$. Then the following are *sufficient conditions* for $\{x_t\}_{t=0}^{\infty}$ to be weakly stationary process:

(i) μ_0 is the eigenvector associated to the eigenvalue $\lambda_1 = 1$ of A:

$$\mu_0' = \mu_0' A.$$

(ii) All other eigenvalues of A are smaller than 1 in absolute value:

$$|\lambda_i| < 1 \quad i = 2,\, ...,\, n.$$

To see this, notice that condition (i) implies that

$$x_{t+1} - \mu_0 = A(x_t - \mu_0) + C w_{t+1}.$$

Then,

$$\Gamma_0 = \Gamma(0) \equiv E\left[(x_t - \mu_0)(x_t - \mu_0)'\right] = A\Gamma(0)A' + CC'$$

and

$$\Gamma(k) \equiv E\left[(x_{t+k} - \mu_0)(x_t - \mu_0)'\right] = A^k \Gamma(0).$$

This is the matrix version of the autocovariance function $\gamma(t,\, k)$ presented above. Notice that we drop t as a variable in this function.

Example 6.6 *Let* $x_t = y_t \in \Re$*,* $A = \rho$*,* $C = \sigma^2$*, and* $w_t = \dfrac{\varepsilon_t}{\sigma}$ *- we are accommodating the AR(1) process seen before to this notation. We can do the following change of variables:*

$$\begin{aligned} \widehat{y}_t &= \begin{pmatrix} y_t \\ 1 \end{pmatrix} \\ \widehat{y}_{t+1} &= \underbrace{\begin{pmatrix} \rho & b \\ 0 & 1 \end{pmatrix}}_{\widehat{A}} \widehat{y}_t + \begin{pmatrix} \sigma \\ 0 \end{pmatrix} w_{t+1}. \end{aligned}$$

Then, using the previous results and ignoring the constant, we get

$$\begin{aligned} \Gamma(0) &= \rho^2 \Gamma(0) + \sigma^2 \\ &\Rightarrow \Gamma(0) = \frac{\sigma^2}{1-\rho^2}. \end{aligned}$$

6.2 Maximization under uncertainty

We will approach this topic by illustrating with examples. Let us begin with a simple 2-period model, where an agent faces a decision problem in which he needs to make the following choices:

1. Consume and save in period 0.
2. Consume and work in period 1.

The uncertainty arises in the income of period 1 through the stochasticity of the wage. We will assume that there are n possible states of the world in period 1, i.e.

$$\omega^2 \in \{\omega_1, ..., \omega_n\},$$

where $\pi_i \equiv \Pr[\omega^2 = \omega_i]$, for $i = 1, ..., n$.

The consumer's utility function has the von Neumann-Morgenstern type, i.e. he is an expected utility maximizer. Leisure in the second period is valued:

$$U = \sum_{i=1}^{n} \pi_i u(c_0, c_{1i}, n_i) \equiv E[u(c_0, c_{1i}, n_i)].$$

Specifically, the utility function is assumed to have the form

$$U = u(c_0) + \beta \sum_{i=1}^{n} \pi_i [u(c_{1i}) + v(n_i)],$$

where $v'(n_i) < 0$.

Market structure: incomplete markets

We will assume that there is a "risk free" asset denoted by a, and priced q, such that every unit of a purchased in period 0 pays 1 unit in period 1, whatever the state of the world. The consumer faces the following budget restriction in the first period:

$$c_0 + aq = I.$$

At each realization of the random state of the world, his budget is given by

$$c_{1i} = a + w_i n_i \quad i = 1, ..., n.$$

The consumer's problem is therefore

$$\max_{c_0,\, a,\, \{c_{1i},\, n_{1i}\}_{i=1}^{n}} u(c_0) + \beta \sum_{i=1}^{n} \pi_i [u(c_{1i}) + v(n_i)]$$

$$\text{s.t.} \quad c_0 + aq = I$$
$$c_{1i} = a + w_i n_i, \; i = 1, ..., n.$$

The first-order conditions are

$$c_0 \; : \; u'(c_0) = \lambda = \sum_{i=1}^{n} \lambda_i R,$$

where $R \equiv \frac{1}{q}$,

$$\begin{aligned} c_{1i} \; &: \; \beta\pi_i u'(c_{1i}) = \lambda_i \\ n_{1i} \; &: \; -\beta\pi_i v'(n_{1i}) = \lambda_i w_i \\ &\Rightarrow -u'(c_{1i})\, w_i = v'(n_{1i}) \end{aligned}$$

$$\begin{aligned} u'(c_0) &= \beta \sum_{i=1}^{n} \pi_i u'(c_{1i})\, R \\ &\equiv \beta E\left[u'(c_{1i})\, R\right]. \end{aligned}$$

The interpretation of the last expression is both straightforward and intuitive: on the margin, the consumer's marginal utility from consumption at period 0 is equated to the *discounted expected* marginal utility from consuming R units in period 1.

Example 6.7 *Let $u(c)$ belong to the CIES class; that is $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$. This is a common assumption in the literature. Recall that σ is the coefficient of relative risk aversion (the higher σ, the less variability in consumption across states the consumer is willing to suffer) and its inverse is the elasticity of intertemporal substitution (the higher σ, the less willing the consumer is to experience the fluctuations of consumption over time). In particular, let $\sigma = 1$, then $u(c) = \log(c)$. Assume also that $v(n) = \log(1-n)$. Replacing in the first-order conditions, these assumptions yield*

$$c_{1i} = w_i\,(1 - n_i)$$

and using the budget constraint at i, we get

$$c_{1i} = \frac{a + w_i}{2}.$$

Therefore,

$$\frac{q}{I - aq} = \beta \sum_{i=1}^{n} \pi_i \frac{2}{a + w_i}.$$

From this equation we get a unique solution, even if not explicit, for the amount of savings given the price q. Finally, notice that we do not have complete insurance in this model (why?).

Market structure: complete markets

We will now modify the market structure in the previous example. Instead of a risk free asset yielding the same payout in each state, we will allow for "Arrow securities" (state-contingent claims): n assets are traded in period 0, and each unit of asset i purchased pays off 1 unit if the realized state is i, and 0 otherwise. The new budget constraint in period 0 is

$$c_0 + \sum_{i=1}^{n} q_i a_i = I.$$

In the second period, if the realized state is i then the consumer's budget constraint is:

$$c_{1i} = a_i + n_i w_i.$$

Notice that a risk free asset can be constructed by purchasing one unit of each a_i. Assume that the total price paid for such a portfolio is the same as before, i.e.

$$q = \sum_{i=1}^{n} q_i.$$

The question is whether the consumer will be better or worse off with this market structure than before. Intuitively, we can see that the structure of wealth transfer across periods that was available before (namely, the risk free asset) is also available now at the same cost. Therefore, the agent could not be worse off. Moreover, the market structure now allows the wealth transfer across periods to be state-specific: not only can the consumer reallocate his income between periods 0 and 1, but also move his wealth across states of the world. Conceptually, this added ability to move income across states will lead to a welfare improvement if the w_i's are nontrivially random, and if preferences show risk aversion (i.e. if the utility index $u(\cdot)$ is strictly concave).

Solving for a_i in the period-1 budget constraints and replacing in the period-0 constraint, we get

$$c_0 + \sum_{i=1}^{n} q_i c_{1i} = I + \sum_{i=1}^{n} q_i w_i n_i.$$

We can interpret this expression in the following way: q_i is the price, in terms of c_0, of consumption goods in period 1 if the realized state is i; $q_i w_i$ is the remuneration to labor if the realized state is i, measured in term of c_0 (remember that budget consolidation only makes sense if all expenditures and income are measured in the same unit of account (in this case it is a monetary unit), where the price of c_0 has been normalized to 1, and q_i is the resulting level of relative prices).

Notice that we have thus reduced the $n + 1$ constraints to 1, whereas in the previous problem we could only eliminate one and reduce them to n. This budget consolidation is a consequence of the free reallocation of wealth across states.

The first-order conditions are

$$c_0 \ : \ u'(c_0) = \lambda$$

$$c_{1i} \ : \ \beta\pi_i u'(c_{1i}) = q_i\lambda$$

$$n_{1i} \ : \ \beta\pi_i v'(n_i) = -q_i w_i \lambda$$

$$\Rightarrow -u'(c_{1i})\, w_i = v'(n_{1i})$$

$$u'(c_0) = \frac{\beta\pi_i}{q_i} u'(c_{1i}), \ i = 1, ...n.$$

The first condition (intra-state consumption-leisure choice) is the same as with incomplete markets. The second condition reflects the added flexibility in allocation of consumption: the agent now not only makes consumption-saving decision in period 0, but also chooses consumption pattern across states of the world.

Under this equilibrium allocation the marginal rates of substitution between consumption in period 0 and consumption in period 1, for any realization of the state of the world, is given by

$$MRS(c_0, c_{1i}) = q_i,$$

and the marginal rates of substitution across states are

$$MRS(c_{1i}, c_{1j}) = \frac{q_i}{q_j}.$$

Example 6.8 *Using the utility function from the previous example, the first-order conditions (together with consolidated budget constraint) can be rewritten as*

$$c_0 = \frac{1}{1+2\beta}\left(I + \sum_{i=1}^{n} q_i w_i\right)$$

$$c_{1i} = \beta c_0 \frac{\pi_i}{q_i}$$

$$n_i = 1 - \frac{c_{1i}}{w_i}.$$

The second condition says that consumption in each period is proportional to consumption in c_0. This proportionality is a function of the cost of insurance: the higher q_i in relation to π_i, the lower the wealth transfer into state i.

6.2.1 Stochastic neoclassical growth model

Notation

We introduce uncertainty into the neoclassical growth model through a stochastic shock affecting factor productivity. A very usual assumption is that of a *neutral* shock, affecting total factor productivity (TFP). Under certain assumptions (for example, Cobb-Douglas $y = AK^{\alpha}n^{1-\alpha}$ production technology), a productivity shock is always *neutral*, even if it is modelled as affecting a specific component (capital K, labor n, technology A).

Specifically, a neoclassical (constant returns to scale) aggregate production function subject to a TFP shock has the form

$$F_t\left(k_t,\, 1\right) = z_t f\left(k_t\right),$$

where z is a stochastic process, and the realizations z_t are drawn from a set Z: $z_t \in Z$, $\forall t$. Let Z^t denote a t-times Cartesian product of Z. We will assume throughout that Z is a countable set (a generalization of this assumption only requires to generalize the summations into integration - however this brings in additional technical complexities which are beyond the scope of this course).

Let z^t denote a *history* of realizations: a t-component vector keeping track of the previous values taken by the z_j for all periods j from 0 to t:

$$z^t = \left(z_t,\, z_{t-1},\, ...,\, z_0\right).$$

Notice that $z^0 = z_0$, and we can write $z^t = (z_t,\, z^{t-1})$.

Let $\pi\left(z^t\right)$ denote the probability of occurrence of the event $(z_t,\, z_{t-1},\, ...,\, z_0)$. Under this notation, a first order Markov process has

$$\pi\left[\left(z_{t+1},\, z^t\right)\left|z^t\right.\right] = \pi\left[\left(z_{t+1},\, z_t\right)\left|z_t\right.\right]$$

(care must be taken of the objects to which probability is assigned).

Sequential formulation

The planning problem in sequential form in this economy requires to maximize the function

$$\sum_{t=0}^{\infty}\sum_{z^t\in Z^t}\beta^t\pi\left(z^t\right)u\left[c_t\left(z^t\right)\right] \equiv E\left[\sum_{t=0}^{\infty}\beta^t u\left[c_t\right]\right].$$

Notice that as t increases, the dimension of the space of events Z^t increases. The choice variables in this problem are the consumption and investment amounts at each date and for each possible realization of the sequence of shocks as of that date. The consumer has to choose a stochastic process for c_t and another one for k_{t+1}:

$$\begin{aligned} &c_t\left(z^t\right) \quad \forall z^t,\ \forall t \\ &k_{t+1}\left(z^t\right) \quad \forall z^t,\ \forall t. \end{aligned}$$

Notice that now there is only one kind of asset (k_{t+1}) available at each date.

Let $(t,\, z^t)$ denote a realization of the sequence of shocks z^t as of date t. The budget constraint in this problem requires that the consumer chooses a consumption and investment amount that is feasible at each $(t,\, z^t)$:

$$c_t\left(z^t\right) + k_{t+1}\left(z^t\right) \leq z_t f\left[k_t\left(z^{t-1}\right)\right] + (1-\delta)\, k_t\left(z^{t-1}\right).$$

You may observe that this restriction is consistent with the fact that the agent's information at the moment of choosing is z^t.

Assuming that the utility index $u(\cdot)$ is strictly increasing, we may as well write the restriction in terms of equality. Then the consumer solves

$$\max_{\{c_t(z^t),\, k_{t+1}(z^t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} \beta^t \pi\left(z^t\right) u\left[c_t\left(z^t\right)\right] \tag{6.1}$$

$$\begin{aligned} \text{s.t.} \quad & c_t\left(z^t\right) + k_{t+1}\left(z^t\right) = z_t f\left[k_t\left(z^{t-1}\right)\right] + (1-\delta)\, k_t\left(z^{t-1}\right), \ \forall\, (t,\, z^t) \\ & k_0 \text{ given.} \end{aligned}$$

Substituting the expression for $c_t\left(z^t\right)$ from budget constraint, the first-order condition with respect to $k_{t+1}\left(z^t\right)$ is

$$-\pi\left(z^t\right) u'\left[c_t\left(z^t\right)\right] + \sum_{z^{t+1} \in Z^{t+1}} \beta \pi\left(z_{t+1},\, z^t\right) u'\left[c_{t+1}\left(z_{t+1},\, z^t\right)\right] \times$$

$$\times \left[z_{t+1} f'\left[k_{t+1}\left(z^t\right)\right] + 1 - \delta\right] = 0.$$

Alternatively, if we denote $\pi\left[\left(z_{t+1},\, z^t\right) \left| z^t \right.\right] \equiv \frac{\pi\left(z_{t+1},\, z^t\right)}{\pi\left(z^t\right)}$, then we can write

$$\begin{aligned} u'\left[c_t\left(z^t\right)\right] = & \sum_{z^{t+1} \in Z^{t+1}} \beta \pi\left[\left(z_{t+1},\, z^t\right) \left| z^t \right.\right] u'\left[c_{t+1}\left(z_{t+1},\, z^t\right)\right] \times \\ & \times \left[z_{t+1} f'\left[k_{t+1}\left(z^t\right)\right] + 1 - \delta\right], \\ \equiv\; & E_{z^t}\left[u'\left[c_{t+1}\left(z_{t+1},\, z^t\right)\right] R_{t+1}\right], \end{aligned} \tag{6.2}$$

where $R_{t+1} \equiv z_{t+1} f'\left[k_{t+1}\left(z^t\right)\right] + 1 - \delta$ is the marginal return on capital realized for each z_{t+1}.

(6.2) is a nonlinear, stochastic difference equation. In general, we will not be able to solve it analytically, so numerical methods or linearization techniques will be necessary.

Recursive formulation

The planner's problem in recursive version is

$$V(k,\, z) = \max_{k'} \left\{ u\left[z f(k) - k' + (1-\delta)\, k\right] + \beta \sum_{z' \in Z} \pi\left(z' \left| z \right.\right) V\left(k',\, z'\right) \right\}, \tag{6.3}$$

where we have used a first order Markov assumption on the process $\{z_t\}_{t=0}^{\infty}$. The solution to this problem involves the policy rule

$$k' = g(k,\, z).$$

If we additionally assume that Z is not only countable but finite, i.e.

$$Z = \{z_1,\, ...,\, z_n\},$$

then the problem can also be written as

$$V_i(k) = \max_{k'} \left\{ u\left[z_i f(k) - k' + (1-\delta) k\right] + \beta \sum_{j=1}^{n} \pi_{ij} V_j(k') \right\},$$

where π_{ij} denotes the probability of moving from state i into state j, i.e.

$$\pi_{ij} \equiv \pi\left[z_{t+1} = z_j \left| z_t = z_i \right.\right].$$

Stationary stochastic process for (k,z)

Let us suppose that we have $g(k,z)$ (we will show later how to obtain it by linearization). What we are interested is what will happen in the long run. We are looking for what is called stationary process for (k,z), i.e. probability distribution over values of (k,z), which is preserved at $t+1$ if applied at time t. It is analogous to the stationary (or invariant) distribution of a Markov process.

Example 6.9 *Let us have a look at a simplified stochastic version, where the shock variable z takes on only two values:*

$$z \in \{z_l, z_h\}.$$

An example of this kind of process is graphically represented in Figure 6.1.

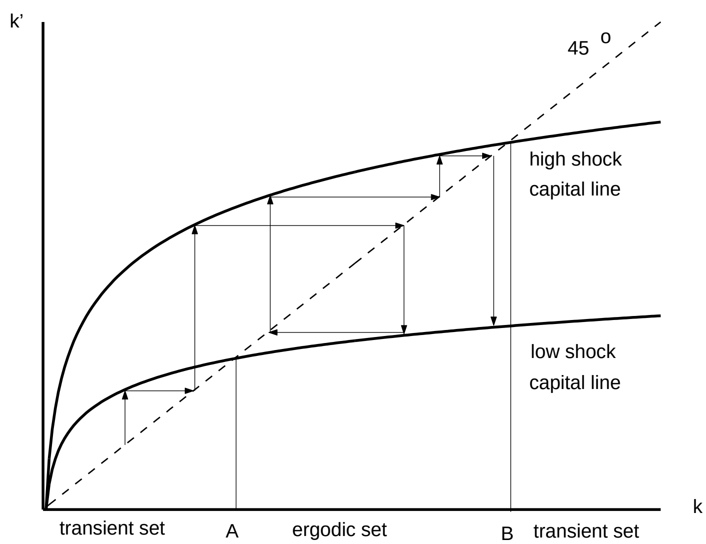

Figure 6.1: An example of (k,z) stochastic process when $z \in \{z_l, z_h\}$

*Following the set-up, we get two sets of possible values of capital, which are of significance for stationary stochastic distribution of (k,z). The first one is the **transient set**, which denotes a set of values of capital, which cannot occur in the long run. It is depicted in Figure 6.1. The probability of leaving the transient set is equal to the probability of capital reaching* *a value higher or equal to A, which is possible only with a high shock. This probability is non-zero and the capital will therefore get beyond A at least once in the long run. Thereafter, the capital will be in the* ***ergodic set****, which is a set, that the capital will never leave once it is there. Clearly, the interval between A and B is an ergodic set since there is no value of capital from this interval and a shock which would cause the capital to take a value outside of this interval in the next period. Also, there is a transient set to the right of B.*

Let $P(k,z)$ denote the joint density, which is preserved over time. As the stochastic process has only two possible states, it can be represented by the density function $P(k,z) = (P_h(k), P_l(k))$. From the above discussion, it is clear to see that the density will be non-zero only for those values of capital that are in the ergodic set. The following are the required properties of $P(k,z)$:

1. *Prob* $[k \leq \bar{k}, z = z_h] = \int_{k \leq \bar{k}} P_h(k)dk =$
   $= \left[\int_{k:g_h(k) \leq \bar{k}} P_h(k)dk\right] \pi_{hh} + \left[\int_{k:g_l(k) \leq \bar{k}} P_l(k)dk\right] \pi_{lh}$

2. *Prob* $[k \leq \bar{k}, z = z_l] = \int_{k \leq \bar{k}} P_l(k)dk =$
   $= \left[\int_{k:g_h(k) \leq \bar{k}} P_h(k)dk\right] \pi_{hl} + \left[\int_{k:g_l(k) \leq \bar{k}} P_l(k)dk\right] \pi_{ll}.$

Note that the above conditions imply that

1. $\int (P_h(k) + P_l(k))dk = 1$ *and*

2. $\int P_h(k)dk = \pi_h$
   $\int P_l(k)dk = \pi_l,$

where π_l and π_h are invariant probabilities of the low and high states.

Solving the model: linearization of the Euler equation

Both the recursive and the sequential formulation lead to the Stochastic Euler Equation

$$u'(c_t) = \beta E_{z_t}\left[u'(c_{t+1})\left[z_{t+1}f'(k_{t+1}) + 1 - \delta\right]\right]. \tag{6.4}$$

Our strategy to solve this equation will be to use a linear approximation of it around the deterministic steady state. We will guess a linear policy function, and replace the choice variables with it. Finally, we will solve for the coefficients of this linear guess.

We rewrite (6.4) in terms of capital and using dynamic programming notation, we get

$$\begin{aligned} u'\left[zf(k) + (1-\delta)k - k'\right] = \beta E_z\left[u'\left[z'f(k') + (1-\delta)k' - k''\right] \times \right. \\ \left. \times \left[z'f'(k') + 1 - \delta\right]\right]. \end{aligned} \tag{6.5}$$

Denote

$$\begin{aligned} LHS &\equiv u'\left[zf(k) + (1-\delta)\,k - k'\right] \\ RHS &\equiv \beta E_z\left[u'\left[z'f(k') + (1-\delta)\,k' - k''\right]\left[z'f'\left(k'\right) + 1 - \delta\right]\right]. \end{aligned}$$

Let $\overline{k}$ be the steady state associated with the realization $\{z_t\}_{t=0}^{\infty}$ that has $z_t = \overline{z}$ for all but a finite number of periods t. That is, $\overline{z}$ is the long run value of z.

Example 6.10 *Suppose that $\{z_t\}_{t=0}^{\infty}$ follows an AR(1) process*

$$z_{t+1} = \rho z_t + (1-\rho)\,\overline{z} + \varepsilon_{t+1},$$

where $|\rho| < 1$. If $E\left[\varepsilon_t\right] = 0$, $E\left[\varepsilon_t^2\right] = \sigma^2 < \infty$, and $E\left[\varepsilon_t \varepsilon_{t+j}\right] = 0 \quad \forall j \geq 1$, then by the Law of Large Numbers we get that

$$plim\ z_t = \overline{z}.$$

Having the long run value of z_t, the associated steady state level of capital $\overline{k}$ is solved from the usual deterministic Euler equation:

$$\begin{aligned} u'\left(\overline{c}\right) &= \beta u'\left(\overline{c}\right)\left[\overline{z}f(\overline{k}) + 1 - \delta\right] \\ \Rightarrow \frac{1}{\beta} &= \overline{z}f(\overline{k}) + 1 - \delta \\ \Rightarrow \overline{k} &= f^{-1}\left(\frac{\beta^{-1} - (1-\delta)}{\overline{z}}\right) \\ \Rightarrow \overline{c} &= \overline{z}f(\overline{k}) - \delta\overline{k}. \end{aligned}$$

Let

$$\begin{aligned} \widehat{k} &\equiv k - \overline{k} \\ \widehat{z} &\equiv z - \overline{z} \end{aligned}$$

denote the variables expressed as deviations from their steady state values. Using this notation we write down a first order Taylor expansion of (6.5) around the long run values as

$$\begin{aligned} LHS &\approx LLHS \equiv a_L\widehat{z} + b_L\widehat{k} + c_L\widehat{k}' + d_L \\ RHS &\approx LRHS \equiv E_z\left[a_R\widehat{z}' + b_R\widehat{k}' + c_R\widehat{k}''\right] + d_R, \end{aligned}$$

where the coefficients a_L, a_R, b_L, etc. are the derivatives of the expressions LHS and RHS with respect to the corresponding variables, evaluated at the steady state (for example, $a_L = u''(\overline{c})f(\overline{k})$). In addition, $LLHS = LRHS$ needs to hold for $\widehat{z} = \widehat{z}' = \widehat{k} = \widehat{k}' = \widehat{k}'' = 0$ (the steady state), and therefore $d_L = d_R$.

Next, we introduce our linear policy function guess in terms of deviations with respect to the steady state as

$$\widehat{k}' = g_k\widehat{k} + g_z\widehat{z}.$$

The coefficients g_k, g_z are unknown. Substituting this guess into the linearized stochastic Euler equation, we get

$$\begin{aligned}
LLHS &= a_L\widehat{z} + b_L\widehat{k} + c_L g_k\widehat{k} + c_L g_z\widehat{z} + d_L \\
LRHS &= E_z\left[a_R\widehat{z}' + b_R g_k\widehat{k} + b_R g_z\widehat{z} + c_R g_k\widehat{k}' + c_R g_z\widehat{z}'\right] + d_R \\
&= E_z\Big[a_R\widehat{z}' + b_R g_k\widehat{k} + b_R g_z\widehat{z} + c_R g_k^2\widehat{k} + c_R g_k g_z\widehat{z} + \\
&\quad + c_R g_z\widehat{z}'\Big] + d_R \\
&= a_R E_z\left[\widehat{z}'\right] + b_R g_k\widehat{k} + b_R g_z\widehat{z} + c_R g_k^2\widehat{k} + c_R g_k g_z\widehat{z} + \\
&\quad + c_R g_z E_z\left[\widehat{z}'\right] + d_R
\end{aligned}$$

and our equation is

$$LLHS = LRHS. \tag{6.6}$$

Notice that d_L, d_R will simplify away. Using the assumed form of the stochastic process $\{z_t\}_{t=0}^{\infty}$, we can replace $E_z\left[\widehat{z}'\right]$ by $\rho\widehat{z}$.

The system (6.6) needs to hold for all values of $\widehat{k}$ and $\widehat{z}$. Given the values of the coefficients a_i, b_i, c_i (for $i = L, R$), the task is to find the values of g_k, g_z that solve the system. Rearranging, (6.6) can be written as

$$\widehat{z}A + E_z\left[\widehat{z}'\right]B + \widehat{k}C = 0,$$

where

$$\begin{aligned}
A &= a_L + c_L g_z - b_R g_z - c_R g_k g_z \\
B &= -a_R - c_R g_z \\
C &= b_L + c_L g_k - b_R g_k - c_R g_k^2.
\end{aligned}$$

As C is a second order polynomial in g_k, the solution will involve two roots. We know that the value smaller than one in absolute value will be the stable solution to the system.

Example 6.11 *Let $\{z_t\}_{t=0}^{\infty}$ follow an AR(1) process, as in the previous example:*

$$z_{t+1} = \rho z_t + (1-\rho)\,\overline{z} + \varepsilon_{t+1}.$$

Then,

$$\begin{aligned}
\widehat{z}' &\equiv z' - \overline{z} \\
&= \rho z + (1-\rho)\,\overline{z} + \varepsilon' - \overline{z} \\
&= \rho\left(z - \overline{z}\right) + \varepsilon'.
\end{aligned}$$

It follows that

$$E_z\left[\widehat{z}'\right] = \rho\widehat{z},$$

and

$$LRHS = a_R\rho\widehat{z} + b_R g_k\widehat{k} + b_R g_z\widehat{z} + c_R g_k^2\widehat{k} + c_R g_k g_z\widehat{z} + c_R g_z\rho\widehat{z} + d_R$$

We can rearrange (6.6) to

$$\widehat{z}A + \widehat{k}B = 0,$$

where

$$\begin{aligned} A &= a_L + c_L g_z - a_R\rho - b_R g_z - c_R g_k g_z - c_R g_z\rho \\ B &= b_L + c_L g_k - b_R g_k - c_R g_k^2. \end{aligned}$$

The solution to (6.6) requires

$$\begin{aligned} A &= 0 \\ B &= 0. \end{aligned}$$

Therefore, the procedure is to solve first for g_k from B (picking the value less than one) and then use this value to solve for g_z from A.

Simulation and impulse response

Once we have solved for the coefficients g_k, g_z, we can *simulate* the model by drawing values of $\{\widehat{z}_t\}_{t=0}^{T}$ from the assumed distribution, and an arbitrary $\widehat{k}_0$. This will yield a stochastic path for capital from the policy rule

$$\widehat{k}_{t+1} = g_k\widehat{k}_t + g_z\widehat{z}_t.$$

We may also be interested in observing the effect on the capital accumulation path in an economy if there is a one-time productivity shock $\widehat{z}$, which is the essence of impulse response. The usual procedure for this analysis is to set $\widehat{k}_0 = 0$ (that is, we begin from the steady state capital stock associated with the long run value $\overline{z}$) and $\widehat{z}_0$ to some arbitrary number. The values of $\widehat{z}_t$ for $t > 0$ are then derived by eliminating the stochastic component in the $\{\widehat{z}_t\}_{t=0}^{T}$ process.

For example, let $\{z_t\}_{t=0}^{\infty}$ be an AR(1) process as in the previous examples, then:

$$\widehat{z}_{t+1} = \rho\widehat{z}_t + \varepsilon_t.$$

Let $\widehat{z}_0 = \Delta$, and set $\varepsilon_t = 0$ for all t. Using the policy function, we obtain the following path for capital:

$$\begin{aligned} \widehat{k}_0 &= 0 \\ \widehat{k}_1 &= g_z\Delta \\ \widehat{k}_2 &= g_k g_z\Delta + g_z\rho\Delta = (g_k g_z + g_z\rho)\,\Delta \\ \widehat{k}_3 &= \left(g_k^2 g_z + g_k g_z\rho + g_z\rho^2\right)\Delta \\ \vdots &= \vdots \\ \widehat{k}_t &= \left(g_k^{t-1} + g_k^{t-2}\rho + \ \dots\ + g_k\rho^{t-2} + \rho^{t-1}\right) g_z\Delta \end{aligned}$$

and

$$|g_k| < 1 \text{ \& } |\rho| < 1 \ \Rightarrow \lim_{t\to\infty} \widehat{k}_t = 0.$$

The capital stock converges back to its steady state value if $|g_k| < 1$ and $|\rho| < 1$.

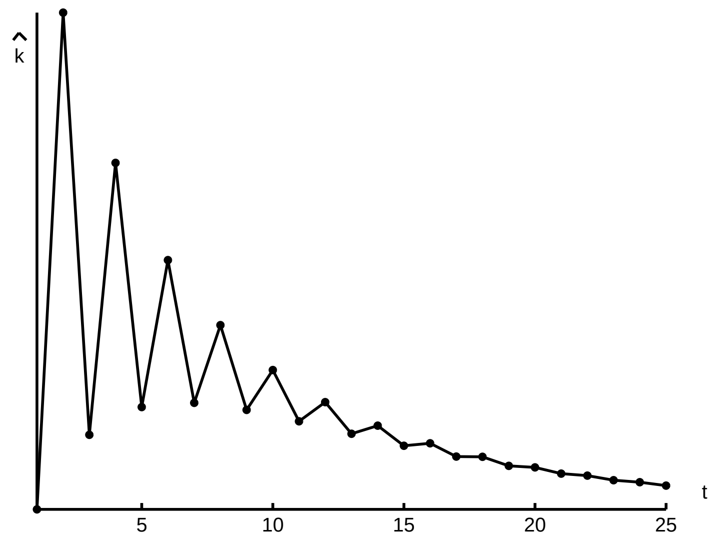

Figure 6.2: An example of an *impulse response* plot, using $g_z = 0.8$, $g_k = 0.9$, $\rho = -0.75$

References and comments on the linear-quadratic setup

You can find most of the material we have discussed on the neoclassical growth model in King, Plosser and Rebelo (1988). Hansen and Sargent (1988) discuss the model in a linear-quadratic environment, which assumes that the production technology is linear in z and k, and u is quadratic:

$$\begin{aligned} y(z,\, k) &= a_y z + b_y k \\ u(c) &= -a_u (c - c_u)^2 + b_u. \end{aligned}$$

This set-up leads to a linear Euler equation, and therefore the linear policy function guess is exact. In addition, the linear-quadratic model has a property called "certainty equivalence", which means that g_k and g_z do not depend on second or higher order moments of the shock ε and it is possible to solve the problem, at all t, by replacing z_{t+k} with $E_t\left[z_{t+k}\right]$ and thus transform it into a deterministic problem.

This approach provides an alternative to linearizing the stochastic Euler equation. We can solve the problem by replacing the return function with a quadratic approximation, and the (technological) constraint by a linear function. Then we solve the resulting linear-quadratic problem

$$\sum_{t=0}^{\infty} \beta^t \underbrace{u\left[F\left(k_t\right)+(1-\delta) k_t-k_{t+1}\right]}_{\text{Return function}}.$$

The approximation of the return function can be done by taking a second order Taylor series expansion around the steady state. This will yield the same results as the linearization.

Finally, the following shortfalls of the linear-quadratic setup must be kept in mind:

- The quadratic return function leads to satiation: there will be a consumption level with zero marginal utility.

- Non-negativity constraints may cause problems. In practice, the method requires such constraints not to bind. Otherwise, the Euler equation will involve Lagrange multipliers, for a significant increase in the complexity of the solution.

- A linear production function implies a constant-marginal-product technology, which may not be consistent with economic intuition.

Recursive formulation issue

There is one more issue to discuss in this section and it involves the choice of state variable in recursive formulation. Let us consider the following problem of the consumer:

$$\max_{\{c_t(z^t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} \beta^t \pi(z^t) u(c_t(z^t))$$

$$\begin{aligned}
\text{s.t.} \quad & z^t=(z_l, z^{t-1}): \ c_t(z^t)+q_{h,t}(z^t)a_{h,t+1}(z^t)+q_{l,t}(z^t)a_{l,t+1}(z^t)=\omega_t(z^t)+a_{l,t}(z^{t-1}) \\
& z^t=(z_h, z^{t-1}): \ c_t(z^t)+q_{h,t}(z^t)a_{h,t+1}(z^t)+q_{l,t}(z^t)a_{l,t+1}(z^t)=\omega_t(z^t)+a_{h,t}(z^{t-1}), \\
& \text{both constraints } \forall t, \ \forall z^t \text{ and no-Ponzi-game condition,}
\end{aligned}$$

where z_t follows a first order Markov process and even more specifically, we only have two states, i.e. $z_t \in \{z_h, z_l\}$. As can be seen, we have two budget constraints, depending on the state at time t.

Let us now consider the recursive formulation of the above-given problem. To simplify matters, suppose that

$$\begin{aligned}
z_t = z_l: \ \omega_t(z^t) &= \omega_l \\
z_t = z_h: \ \omega_t(z^t) &= \omega_h.
\end{aligned}$$

What are our state variables going to be? Clearly, z_t has to be one of our state variables. The other will be wealth w (differentiate from the endowment ω), which we can define as a sum of the endowment and the income from asset holdings:

$$\begin{aligned}
z_t = z_l: \ w_t(z^t) &= \omega_l + a_{l,t}(z^{t-1}) \\
z_t = z_h: \ w_t(z^t) &= \omega_h + a_{h,t}(z^{t-1}).
\end{aligned}$$

The recursive formulation is now

$$V(w, z_i) \equiv V_i(w) =$$

$$= \max_{a_l', a_h'} \left\{ u(w - q_{ih}a_h' - q_{il}a_l') + \beta \left[\pi_{ih} V_h(\underbrace{\omega_h + a_h'}_{w_h'}) + \pi_{il} V_l(\underbrace{\omega_l + a_l'}_{w_l'}) \right] \right\},$$

where the policy rules are now

$$\begin{aligned} a_h' &= g_{ih}(w) \\ a_l' &= g_{il}(w), \; i = l, h. \end{aligned}$$

Could we use a as a state variable instead of w? Yes, we could, but that would actually imply two state variables - a_h and a_l. Since the state variable is to be a variable which expresses the relevant information as succinctly as possible, it is w that we should use.

6.3 Competitive equilibrium under uncertainty

The welfare properties of competitive equilibrium are affected by the introduction of uncertainty through the *market structure*. The relevant distinction is whether such structure involves *complete* or *incomplete markets*. Intuitively, a *complete markets* structure allows trading in each single commodity. Recall our previous discussion of the neoclassical growth model under uncertainty where commodities are defined as consumption goods indexed by time and state of the world. For example, if z_1^t and z_2^t denote two different realizations of the random sequence $\{z_j\}_{j=0}^t$, then a unit of the physical good c consumed in period t if the state of the world is z_1^t (denoted by $c_t(z_1^t)$) is a commodity different from $c_t(z_2^t)$. A complete markets structure will allow contracts between parties to specify the delivery of physical good c in different amounts at $(t,\, z_1^t)$ than at $(t,\, z_2^t)$, and for a different price.

In an incomplete markets structure, such a contract might be impossible to enforce and the parties might be unable to sign a “legal” contract that makes the delivery amount contingent on the realization of the random shock. A usual incomplete markets structure is one where agents may only agree to the delivery of goods on a date basis, regardless of the shock. In short, a contract specifying $c_t(z_1^t) \neq c_t(z_2^t)$ is not enforceable in such an economy.

You may notice that the structure of markets is an assumption of an *institutional* nature and nothing should prevent, in theory, the market structure to be complete. However, markets are incomplete in the real world and this seems to play a key role in the economy (for example in the distribution of wealth, in the business cycle, perhaps even in the *equity premium* puzzle that we will discuss in due time).

Before embarking on the study of the subject, it is worth mentioning that the structure of markets need not be explicit. For example, the accumulation of capital may supply the role of transferring wealth across states of the world (not just across time). But allowing for the transfer of wealth across states is one of the functions specific to markets; therefore, if these are incomplete then capital accumulation can (to some extent) perform this missing function. An extreme example is the deterministic model, in which there is only one state of the world and only transfers of wealth across time are relevant. The possibility of accumulating capital is enough to ensure that markets are complete and allowing agents also to engage in trade of dated commodities is redundant. Another example shows up in real business cycle models, which we shall analyze later on in this course. A usual result in the real business cycle literature (consistent with actual economic data) is that agents choose to accumulate more capital whenever there is a "good" realization of the productivity shock. An intuitive interpretation is that savings play the role of a "buffer" used to smooth out the consumption path, which is a function that markets could perform.

Hence, you may correctly suspect that whenever we talk about market completeness or incompleteness, we are in fact referring not to the actual, explicit contracts that agents are allowed to sign, but to the degree to which they are able to transfer wealth across states of the world. This ability will depend on the institutional framework assumed for the economy.

6.3.1 The neoclassical growth model with complete markets

We will begin by analyzing the neoclassical growth model in an uncertain environment. We assume that, given a stochastic process $\{z_t\}_{t=0}^{\infty}$, there is a market for each consumption commodity $c_t\left(z^t\right)$, as well as for capital and labor services at each date and state of the world. There are two alternative setups: Arrow-Debreu date-0 trading and sequential trading.

Arrow-Debreu date-0 trading

The Arrow-Debreu date-0 competitive equilibrium is

$$\{c_t(z^t), k_{t+1}(z^t), l_t(z^t), p_t(z^t), r_t(z^t), w_t(z^t)\}_{t=0}^{\infty}$$

such that

1. Consumer's problem is to find $\{c_t(z^t), k_{t+1}(z^t), l_t(z^t)\}_{t=0}^{\infty}$ which solve

$$\max_{\{c_t(z^t), k_{t+1}(z^t), l_t(z^t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} \beta^t \pi(z^t) u(c_t(z^t), 1 - l_t(z^t))$$

$$\text{s.t.} \quad \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} p_t\left(z^t\right)\left[c_t\left(z^t\right) + k_{t+1}\left(z^t\right)\right] \leq \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} p_t\left(z^t\right)\left[\left(r_t\left(z^t\right) + 1 - \delta\right) \times \right.$$
$$\left. \times\ k_t\left(z^{t-1}\right) + w_t\left(z^t\right) l_t\left(z^t\right)\right].$$

2. First-order conditions from firm's problem are

$$\begin{aligned} r_t(z^t) &= z_t F_k(k_t(z^{t-1}), l_t(z^t)) \\ w_t(z^t) &= z_t F_l(k_t(z^{t-1}), l_t(z^t)). \end{aligned}$$

3. Market clearing is

$$c_t(z^t) + k_{t+1}(z^t) = (1-\delta)k_t(z^{t-1}) + z^t F(k_t(z^{t-1}), l_t(z^t)), \ \forall t, \ \forall z^t.$$

You should be able to show that the Euler equation in this problem is identical to the Euler equation in the planner's problem.

In this context, it is of interest to mention the so-called no-arbitrage condition, which can be derived from the above-given setup. First, we step inside the budget constraint and retrieve those terms which relate to $k_{t+1}(z^t)$:

- From the LHS: ...$p_t(z^t)k_{t+1}(z^t)$...
- From the RHS: ... $\sum_{z_{t+1}} p_{t+1}(z_{t+1}, z^t) \left[r_{t+1}(z_{t+1}, z^t) + (1-\delta)\right] k_{t+1}(z^t)$...

The no-arbitrage condition is the equality of these two expressions and it says that in equilibrium, the price of a unit of capital must equal the sum of future values of a unit of capital summed across all possible states. Formally, it is

$$k_{t+1}(z^t) \left[p_t(z^t) - \sum_{z_{t+1}} p_{t+1}(z_{t+1}, z^t) \left[r_{t+1}(z_{t+1}, z^t) + (1-\delta)\right] \right] = 0.$$

What would happen if the no-arbitrage condition did not hold? Assuming $k_{t+1}(z^t) \geq 0$, the term in the brackets would have to be non-zero. If this term were greater then zero, we could make infinite "profit" by setting $k_{t+1}(z^t) = -\infty$. Similarly, if the term were less than zero, setting $k_{t+1}(z^t) = \infty$ would do the job. As neither of these can happen in equilibrium, the term in the brackets must equal zero, which means that the no-arbitrage condition must hold in equilibrium.

Sequential trade

In order to allow wealth transfers across dates, agents must be able to borrow and lend. It suffices to have one-period assets, even with an infinite time horizon. We will assume the existence of these one-period assets, and, for simplicity, that Z is a finite set with n possible shock values, as is illustrated in Figure 6.3.

Assume that there are q assets, with asset j paying off r_{ij} consumption units in $t+1$ if the realized state is z_i. The following matrix shows the payoff of each asset for every realization of z_{t+1}:

$$\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_n \end{array} \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & \cdots & a_q \end{array} \\ \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1q} \\ r_{21} & r_{22} & \cdots & r_{2q} \\ r_{31} & r_{32} & \cdots & r_{3q} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nq} \end{pmatrix} \end{array} \equiv R.$$

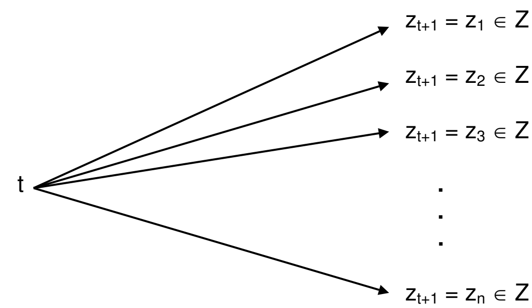

Figure 6.3: The shock z can take n possible values, which belong to Z

Then the portfolio $a = (a_1, a_2, ..., a_q)$ pays p (in terms of consumption goods at $t+1$), where

$$\underbrace{p}_{n\times 1} = \underbrace{R}_{n\times q} \cdot \underbrace{a}_{q\times 1},$$

and each component $p_i = \sum_{j=1}^{q} r_{ij}a_j$ is the amount of consumption goods obtained in state i from holding portfolio a.

What restrictions must we impose on R so that any arbitrary payoff combination $p \in \Re^n$ can be generated (by the appropriate portfolio choice)? Based on matrix algebra, the answer is that we must have

1. $q \geq n$.

2. $rank(R) = n$.

If R satisfies condition number (2) (which presupposes the validity of the first one), then the market structure is *complete*. The whole space $\Re^n$ is *spanned* by R and we say that there is *spanning* .

It is useful to mention Arrow securities which were mentioned before. Arrow security i pays off 1 unit if the realized state is i, and 0 otherwise. If there are $q < n$ different Arrow securities, then the payoff matrix is

$$\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_q \\ \vdots \\ z_n \end{array} \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & \cdots & a_q \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \end{array}.$$

6.3.2 General equilibrium under uncertainty: the case of two agent types in a two-period setting

First, we compare the outcome of the neoclassical growth model with uncertainty and one representative agent with the two different market structures:

- Only (sequential) trade in capital is allowed. There is no *spanning* in this setup as there is only one asset for n states.

- *Spanning* (either with Arrow-Debreu date-0, or sequential trading).

Will equilibria look different with these structures? The answer is no, and the reason is that there is a single agent. Clearly, every loan needs a borrower and a lender, which means that the total borrowing and lending in such an economy will be zero. This translates into the fact that different asset structures do not yield different equilibria.

Let us turn to the case where the economy is populated by more than one agent to analyze the validity of such a result. We will compare the equilibrium allocation of this economy under the market structures (1) and (2) mentioned above.

Assumptions

- Random shock: We assume there are n states of the world corresponding to n different values of the shock to technology to be described as

$$z \in \{z_1,\, z_2,\, ...,\, z_n\}$$

$$\pi_j = \Pr\left[z = z_j\right].$$

  Let $\overline{z}$ denote the expected value of z:

$$\overline{z} = \sum_{j=1}^{n} \pi_j z_j.$$

- Tastes: Agents derive utility from consumption only (not from leisure). Preferences satisfy the axioms of expected utility and are represented by the utility index $u()$. Specifically, we assume that

$$U_i = u_i\left(c_0^i\right) + \beta \sum_{j=1}^{n} \pi_j u_i(c_j^i) \qquad i = 1, 2.$$

  where $u_1(x) = x$, and $u_2(x)$ is strictly concave ($u_2' > 0$, $u_2'' < 0$). We also assume that $\lim_{x \to 0} u_2'(x) = \infty$. In this fashion, agents' preferences exhibit different attitudes towards risk: Agent 1 is *risk neutral* and Agent 2 is *risk averse*.

- Endowments: Each agent is endowed with ω_0 consumption goods in period 0, and with one unit of labor in period 1 (which will be supplied inelastically since leisure is not valued).

- Technology: Consumption goods are produced in period 1 with a constant-returns-to-scale technology represented by the Cobb Douglas production function

$$y_j = z_j K^{\alpha} \left(\frac{n}{2}\right)^{1-\alpha}.$$

  where K, n denote the aggregate supply of capital and labor services in period 1, respectively. We know that $n = 2$, so

$$y_j = z_j K^{\alpha}.$$

  Therefore, the remunerations to factors in period 1, if state j is realized, are given by

$$\begin{aligned} r_j &= z_j \alpha K^{\alpha - 1} \\ w_j &= z_j \frac{(1-\alpha)}{2} K^{\alpha}. \end{aligned}$$

Structure 1 - one asset

Capital is the only asset that is traded in this setup. With K denoting the aggregate capital stock, a_i denotes the capital stock held by agent i, and therefore the asset market clearing requires that

$$a_1 + a_2 = K.$$

The budget constraints for each agent is given by

$$\begin{aligned} c_0^i + a_i &= \omega_0 \\ c_j^i &= a_i r_j + w_j. \end{aligned}$$

To solve this problem, we proceed to maximize each consumer's utility subject to his budget constraint.

Agent 1:

The maximized utility function and the constraints are linear in this case. We therefore use the arbitrage condition to express optimality:

$$\left[-1+\beta\sum_{j=1}^{n}\pi_j r_j\right]a_i=0.$$

For a_i not to be infinite (which would violate the market clearing condition), that part of the arbitrage condition which is in brackets must equal zero. Replacing for r_j, we get then

$$1=\beta\sum_{j=1}^{n}\pi_j\alpha z_j K^{\alpha-1} \tag{6.7}$$

$$\Rightarrow 1=\alpha\beta K^{\alpha-1}\sum_{j=1}^{n}\pi_j z_j.$$

Therefore, the optimal choice of K from Agent 1's preferences is given by

$$K^*=(\overline{z}\alpha\beta)^{\frac{1}{1-\alpha}}.$$

Notice that only the average value of the random shock matters for Agent 1, consistently with this agent being *risk neutral*.

Agent 2:

The Euler equation for Agent 2 is

$$u_2'(\omega_0-a_2)=\beta\sum_{j=1}^{n}\pi_j u_2'\left(a_2r_j^*+w_j^*\right)r_j^*. \tag{6.8}$$

Given K^* from Agent 1's problem, we have the values of r_j^* and w_j^* for each realization j. Therefore, Agent 2's Euler equation (6.8) is one equation in one unknown a_2. Since $\lim_{x\to 0}u_2'(x)=\infty$, there exists a unique solution. Let a_2^* be the solution to (6.8). Then the values of the remaining choice variables are

$$\begin{aligned} a_1^* &= K^*-a_2^* \\ c_0^i &= \omega_0-a_i^*. \end{aligned}$$

More importantly, Agent 2 will face a stochastic consumption prospect for period 1, which is

$$c_j^2=a_2^*r_j^*+w_j^*,$$

where r_j^* and w_j^* are stochastic. This implies that Agent 1 has not provided *full insurance* to Agent 2.

Structure 2 - Arrow securities

It is allowed to trade in n different Arrow securities in this setup. In this case, these securities are (contingent) claims on the total remuneration to capital (you could think of them as rights to collect future dividends in a company, according to the realized state of the world). Notice that this implies *spanning* (i.e. markets are complete). Let a_j denote the Arrow security paying off one unit if the realized state is z_j and zero otherwise. Let q_j denote the price of a_j.

In this economy, agents *save* by accumulating contingent claims (they save by buying future dividends in a company). Total savings are thus given by

$$S \equiv \sum_{j=1}^{n} q_j \left(a_{1j} + a_{2j}\right).$$

Investment is the accumulation of physical capital, K. Then clearing of the savings-investment market requires that:

$$\sum_{j=1}^{n} q_j \left(a_{1j} + a_{2j}\right) = K. \tag{6.9}$$

Constant returns to scale imply that the total remuneration to capital services in state j will be given by $r_j K$ (by Euler Theorem). Therefore, the contingent claims that get activated when this state is realized must exactly match this amount (each unit of "dividends" that the company will pay out must have an owner, but the total claims can not exceed the actual amount of dividends to be paid out).

In other words, clearing of (all of) the Arrow security markets requires that

$$a_{1j} + a_{2j} = K r_j \quad j = 1, ..., n. \tag{6.10}$$

If we multiply both sides of (6.10) by q_j, for each j, and then sum up over j's, we get

$$\sum_{j=1}^{n} q_j \left(a_{1j} + a_{2j}\right) = K \sum_{j=1}^{n} q_j r_j.$$

But, using (6.9) to replace total savings by total investment,

$$K = K \sum_{j=1}^{n} q_j r_j.$$

Therefore the equilibrium condition is that

$$\sum_{j=1}^{n} q_j r_j = 1. \tag{6.11}$$

The equation (6.11) can be interpreted as a no-arbitrage condition, in the following way. The left hand side $\sum_{j=1}^{n} q_j r_j$ is the total price (in terms of foregone consumption units) of the marginal unit of a portfolio yielding the same (expected) marginal return as physical capital investment. And the right hand side is the price (also in consumption units) of a marginal unit of capital investment.

First, suppose that $\sum_{j=1}^{n} q_j r_j > 1$. An agent could in principle make unbounded profits by selling an infinite amount of units of such a portfolio, and using the proceeds from this sale to finance an unbounded physical capital investment. In fact, since no agent would be willing to be on the buy side of such a deal, no trade would actually occur. But there would be an infinite supply of such a portfolio, and an infinite demand of physical capital units. In other words, asset markets would not be in equilibrium. A similar reasoning would lead to the conclusion that $\sum_{j=1}^{n} q_j r_j < 1$ could not be an equilibrium either.

With the equilibrium conditions at hand, we are able to solve the model. With this market structure, the budget constraint of each Agent i is

$$c_0^i + \sum_{j=1}^{n} q_j a_{ij} = \omega_0$$

$$c_j^i = a_{ij} + w_j.$$

Using the first order conditions of Agent 1's problem, the equilibrium prices are

$$q_j = \beta \pi_j.$$

You should also check that

$$K^* = (\overline{z}\alpha\beta)^{\frac{1}{1-\alpha}},$$

as in the previous problem. Therefore, Agent 1 is as well off with the current market structure as in the previous setup.

Agent 2's problem yields the Euler equation

$$u_2'\left(c_0^2\right) = \lambda = q_j^{-1}\beta\pi_j u_2'\left(c_j^2\right).$$

Replacing for the equilibrium prices derived from Agent 1's problem, this simplifies to

$$u_2'\left(c_0^2\right) = u_2'\left(c_j^2\right) \qquad j = 1, ..., n.$$

Therefore, with the new market structure, Agent 2 is able to obtain *full insurance* from Agent 1. From the First Welfare Theorem (which requires completeness of markets) we know that the allocation prevailing under market *Structure 2* is a Pareto optimal allocation. It is your task to determine whether the allocation resulting from *Structure 1* was Pareto optimal as well or not.

6.3.3 General equilibrium under uncertainty: multiple-period model with two agent types

How does the case of infinite number of periods differ from the two-period case? In general, the conclusions are the same and the only difference is the additional complexity added through extending the problem. We shortly summarize both structures. As before, Agent 1 is risk neutral and Agent 2 is risk averse.

Structure 1 - one asset

Agent 1:

Agent 1's problem is

$$\max \sum_{z^t \in Z^t} \sum_{t=0}^{\infty} \beta^t \pi(z^t) c_t(z^t)$$

$$\text{s.t.} \quad c_{1,t}(z^t) + a_{1,t+1}(z^t) = r_t(z^t) a_{1,t}(z^{t-1}) + w_t(z^t).$$

Firm's problem yields (using Cobb-Douglas production function)

$$\begin{aligned} r_t(z^t) &= z_t \alpha k_t^{\alpha-1}(z^{t-1}) + (1-\delta) \\ w_t(z^t) &= z_t \left(\frac{1-\alpha}{2}\right) k_t^{\alpha}(z^{t-1}). \end{aligned}$$

Market clearing condition is

$$a_{1,t+1}(z^t) + a_{2,t+1}(z^t) = k_{t+1}(z^t).$$

First-order condition w.r.t. $a_{1,t+1}(z^t)$ gives us

$$\begin{aligned} 1 &= \beta \sum_{z_{t+1}} \frac{\pi(z_{t+1}, z^t)}{\pi(z^t)} r_{t+1}(z_{t+1}, z^t) \\ \Rightarrow 1 &= \beta E_{z_{t+1}|z^t}(r_{t+1}). \end{aligned}$$

Using the formula for r_{t+1} from firm's first-order conditions, we get

$$\begin{aligned} 1 &= \beta \sum_{z_{t+1}} \pi(z_{t+1}|z^t) \left(z_{t+1} \alpha k_{t+1}^{\alpha-1}(z^t) + (1-\delta) \right) = \\ &= \alpha \beta k_{t+1}^{\alpha-1}(z^t) \underbrace{\sum_{z_{t+1}} \pi(z_{t+1}|z^t) z_{t+1}}_{E(z_{t+1}|z^t)} + \beta(1-\delta) \end{aligned}$$

$$\Rightarrow k_{t+1}(z^t) = \left[\frac{1/\beta - 1 + \delta}{\alpha E(z_{t+1}|z^t)} \right]^{\frac{1}{\alpha-1}}. \tag{6.12}$$

Agent 2:

Agent 2's utility function is $u(c_{2,t}(z^t))$ and his first-order conditions yield

$$u'(c_{2,t}(z^t)) = \beta E_{z_{t+1}|z^t} \left[u'(c_{2,t+1}(z^{t+1}))(1 - \delta + \alpha z_{t+1} k_{t+1}^{\alpha-1}(z^t)) \right].$$

Using the above-given Euler equation and (6.12) together with Agent 2's budget constraint, we can solve for $c_{2,t}(z^t)$ and $a_{2,t+1}(z^t)$. Subsequently, using the market clearing condition gives us the solution for $c_{1,t}(z^t)$.

The conclusion is the same as in the two-period case: Agent 2 does not insure fully and his consumption across states will vary.

Structure 2 - Arrow securities

Agent 1:

The problem is very similar to the one in Structure 1, except for the budget constraint, which is now

$$c_t^1(z^t) + \sum_{j=1}^{n} q_j(z^t) a_{j,t+1}^1(z^t) = a_{i,t}^1(z^{t-1}) + w_t(z^t).$$

As we have more than one asset, the no-arbitrage condition has to hold. It can expressed as

$$\sum_{j=1}^{n} q_j(z^t) a_{j,t+1}(z^t) = k_{t+1}(z^t)$$

$$a_{j,t+1}(z^t) = \left[1 - \delta + r_{t+1}(z_j, z^t)\right] k_{t+1}(z^t)$$

$$\Rightarrow 1 = \sum_{j=1}^{n} q_j(z^t) \left[1 - \delta + r_{t+1}(z_j, z^t)\right].$$

Solving the first-order condition of Agent 1 w.r.t. $a_{j,t+1}^1(z^t)$ yields

$$q_{j,t}(z^t) = \beta \frac{\pi(z_j, z^t)}{\pi(z^t)} = \beta \pi(z_j | z^t), \tag{6.13}$$

which is the formula for prices of the Arrow securities.

Agent 2:

The first-order condition w.r.t. $a_{j,t+1}^2(z^t)$ yields

$$0 = -\beta^t \pi(z^t) q_{j,t}(z^t) u'(c_t^2(z^t)) + \beta^{t+1} \pi(z_j, z^t) u'(c_{t+1}^2(z_j, z^t)).$$

Substituting (6.13) gives us

$$0 = -\beta^t \pi(z^t) \beta \frac{\pi(z_j, z^t)}{\pi(z^t)} u'(c_t^2(z^t)) + \beta^{t+1} \pi(z_j, z^t) u'(c_{t+1}^2(z_j, z^t))$$

$$\Rightarrow u'(c_t^2(z^t)) = u'(c_{t+1}^2(z_j, z^t))$$

$$\Rightarrow c_t^2(z^t) = c_{t+1}^2(z_j, z^t).$$

This result follows from the assumption that $u'(.) > 0$ and $u''(.) < 0$, and yields the same conclusion as in the two-period case, i.e. Agent 2 insures completely and his consumption does not vary across states.

6.3.4 Recursive formulation

The setup is like that studied above: let Agent 1 have a different (say, lower) degree of risk aversion than Agent 2's (though allow more generality, so that Agent 1 is not necessarily risk-neutral). We denote the agent type by superscript and the state of the world by subscript. The stochastic process is a first order Markov process. Using the recursive formulation knowledge from before, we use wealth (denoted by ω to differentiate it from wage, which is denoted by w) as the state variable. More concretely, there are three state variables: individual wealth (ω), the average wealth of risk neutral agents ($\overline{\omega}_1$), and the average wealth of risk averse agents ($\overline{\omega}_2$). The problem of consumer l is then

$$V_i^l(\omega,\overline{\omega}_1,\overline{\omega}_2) = \max_{\{a'_j\}_{j=1}^n} \{u^l(\omega - \sum_j q_{ij}(\overline{\omega}_1,\overline{\omega}_2)a'_j) + \beta \sum_j \pi_{ij} V_j^l[a'_j +$$

$$+ w_j(G_i(\overline{\omega}_1,\overline{\omega}_2)), D_{ij}^1(\overline{\omega}_1,\overline{\omega}_2) + w_j(G_i(\overline{\omega}_1,\overline{\omega}_2)), D_{ij}^2(\overline{\omega}_1,\overline{\omega}_2) + w_j(G_i(\overline{\omega}_1,\overline{\omega}_2))]\}, \qquad (6.14)$$

where

$$D_{ij}^1(\overline{\omega}_1,\overline{\omega}_2) = d_{ij}^1(\overline{\omega}_1,\overline{\omega}_1,\overline{\omega}_2),\ \forall\ i,j,\overline{\omega}_1,\overline{\omega}_2$$
$$D_{ij}^2(\overline{\omega}_1,\overline{\omega}_2) = d_{ij}^2(\overline{\omega}_2,\overline{\omega}_1,\overline{\omega}_2),\ \forall\ i,j,\overline{\omega}_1,\overline{\omega}_2$$
$$G_i(\overline{\omega}_1,\overline{\omega}_2) = \sum_j q_{ij}(\overline{\omega}_1,\overline{\omega}_2)(D_{ij}^1(\overline{\omega}_1,\overline{\omega}_2) + D_{ij}^2(\overline{\omega}_1,\overline{\omega}_2)),\ \forall\ i,\overline{\omega}_1,\overline{\omega}_2.$$

Let $a'_j = d_{ij}^l(\omega,\overline{\omega}_1,\overline{\omega}_2)$ denote the optimal asset choice of the consumer.

From the firm's problem, we get the first-order conditions specifying the wage and the interest rate as

$$\begin{aligned} w_j(k) &= z_j F_l(k,1), \forall\ j,k \\ r_j(k) &= z_j F_k(k,1), \forall\ j,k. \end{aligned}$$

Asseet-market clearing requires that

$$\sum_{l=1}^{2} D_{ij}^l(\overline{\omega}_1,\overline{\omega}_2) \quad = \quad (1-\delta + r_j(G_i(\overline{\omega}_1,\overline{\omega}_2)))G_i(\overline{\omega}_1,\overline{\omega}_2).$$

The formulation is very similar to our previous formulation of recursive competitive equilibrium, with some new unfamiliar notation showing up. Clearly, d_{ij}^l and D_{ij}^l represent individual and aggregate asset choices for agent l, respectively, whereas the capital stock invested for the next period is denoted by G_i. Notice also that capital is not a separate state variable here; what value the capital stock has can, in fact, be backed out from knowledge of i, $\overline{\omega}_1$, and $\overline{\omega}_2$) (how?).

The following are the unknown functions: $V_i^l(.)$, d_{ij}^l, $D_{ij}^l(.)$, $q_{ij}(.)$, $G_i(.)$, $w_j(.)$, $r_j(.)$. It is left as an exercise to identify the elements from the recursive formulation with elements from the sequential formulation.

In the special case where one agent is risk-neutral, it will transpire that $q_{ij}(\overline{\omega}_1,\overline{\omega}_2) = \beta\pi_{ij}$ and that $G_i(\overline{\omega}_1,\overline{\omega}_2) = \left[\frac{1/\beta - 1 + \delta}{\alpha E(z_{t+1}|z_i)}\right]^{\frac{1}{\alpha-1}}$ for all i and $(\overline{\omega}_1,\overline{\omega}_2)$.

6.4 Appendix: basic concepts in stochastic processes

We will introduce the basic elements with which uncertain events are modelled. The main mathematical notion underlying the concept of uncertainty is that of a probability space.

Definition 6.12 *A probability space is a mathematical object consisting of three elements: 1) a set Ω of possible outcomes ω; 2) a collection $\mathcal{F}$ of subsets of Ω that constitute the "events" to which probability is assigned (a σ-algebra); and 3) a set function P that assigns probability values to those events. A probability space is denoted by*

$$(\Omega, \mathcal{F}, P).$$

Definition 6.13 *A σ-algebra ($\mathcal{F}$) is a special kind of family of subsets of a space Ω that satisfy three properties: 1) $\Omega \in \mathcal{F}$, 2) $\mathcal{F}$ is closed under complementation: $E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$, 3) $\mathcal{F}$ is closed under countable union: if $\{E_i\}_{i=1}^{\infty}$ is a sequence of sets such that $E_i \in \mathcal{F}$ $\forall i$, then $(\cup_{i=1}^{\infty} E_i) \in \mathcal{F}$.*

Definition 6.14 *A random variable is a function whose domain is the set of events Ω and whose image is the real numbers (or a subset thereof):*

$$x : \Omega \to \Re.$$

For any real number α, define the set

$$E_\alpha = \{\omega : x(\omega) < \alpha\}.$$

Definition 6.15 *A function x is said to be measurable with respect to the σ-algebra $\mathcal{F}$ (or $\mathcal{F}$-measurable) if the following property is satisfied:*

$$\forall \alpha \in \Re : E_\alpha \in \mathcal{F}.$$

Conceptually, if x is $\mathcal{F}$-measurable then we can assign probability to the event $x < \alpha$ for any real number α. [We may equivalently have used $>$, $\leq$ or $\geq$ for the definition of measurability, but that is beyond the scope of this course. You only need to know that if x is $\mathcal{F}$-measurable, then we can sensibly talk about the probability of x taking values in virtually any subset of the real line you can think of (the *Borel* sets).]

Now define a sequence of σ-algebras as

$$\{\mathcal{F}_t\}_{t=1}^{\infty} : \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq ... \subseteq \mathcal{F}.$$

Conceptually, each σ-algebra $\mathcal{F}_t$ "refines" $\mathcal{F}_{t-1}$, in the sense that distinguishes (in a probabilistic sense) between "more" events than the previous one.

Finally, let a sequence of random variables x_t be $\mathcal{F}_t$-measurable for each t, which models a stochastic process. Consider an $\omega \in \Omega$, and choose an $\alpha \in \Re$. Then for each t, the set $E_{\alpha t} \equiv \{\omega : x_t(\omega) < \alpha\}$ will be a set included in the collection (the σ-algebra) $\mathcal{F}_t$. Since $\mathcal{F}_t \subseteq \mathcal{F}$ for all t, $E_{\alpha t}$ also belongs to $\mathcal{F}$. Hence, we can assign probability to $E_{\alpha t}$ using the set function P and $P[E_{\alpha t}]$ is well defined.

Example 6.16 *Consider the probability space* $(\Omega, \mathcal{F}, P)$*, where*

- $\Omega = [0,\, 1]$.
- $\mathcal{F} = \mathcal{B}$ *(the Borel sets restricted to* $[0,\, 1]$*).*
- $P = \lambda$ *- the length of an interval:* $\lambda([a,\, b]) = b - a$.

Consider the following collections of sets:

$$A_t = \left\{ \left\{ \left[\frac{j}{2^t},\, \frac{j+1}{2^t} \right) \right\}_{j=0}^{2^t-2},\, \left[\frac{2^t-1}{2^t},\, 1 \right] \right\}.$$

For every t*, let* $\mathcal{F}_t$ *be the minimum* σ*-algebra containing* A_t*. Denote by* $\sigma(A_t)$ *the collection of all possible unions of the sets in* A_t *(notice that* $\Omega \in \sigma(A_t)$*). Then* $\mathcal{F}_t = \{\emptyset,\, A_t,\, \sigma(A_t)\}$ *(you should check that this is a* σ*-algebra).*

For example,

$$\begin{gathered} A_1 = \left\{ [0,\, 1],\, \emptyset,\ \left[0,\, \tfrac{1}{2}\right),\, \left[\tfrac{1}{2},\, 1\right] \right\} \\ \Rightarrow \mathcal{F}_1 = \left\{ [0,\, 1],\, \emptyset,\ \left[0,\, \tfrac{1}{2}\right),\, \left[\tfrac{1}{2},\, 1\right] \right\} \\ A_2 = \left\{ \left[0,\, \tfrac{1}{4}\right),\, \left[\tfrac{1}{4},\, \tfrac{1}{2}\right),\, \left[\tfrac{1}{2},\, \tfrac{3}{4}\right),\, \left[\tfrac{3}{4},\, 1\right] \right\} \\ \Rightarrow \sigma(A_2) = \left\{ \left[0,\, \tfrac{1}{2}\right),\, \left[0,\, \tfrac{3}{4}\right),\, \left[\tfrac{1}{4},\, \tfrac{3}{4}\right),\, \left[\tfrac{1}{4},\, 1\right],\, \left[\tfrac{1}{2},\, 1\right],\, \left[0,\, \tfrac{1}{4}\right) \cup \left[\tfrac{1}{2},\, \tfrac{3}{4}\right) \right\} \cup \\ \cup \left\{ \left[0,\, \tfrac{1}{4}\right) \cup \left[\tfrac{1}{2},\, 1\right],\, \left[0,\, \tfrac{1}{4}\right) \cup \left[\tfrac{3}{4},\, 1\right],\, \left[0,\, \tfrac{1}{2}\right) \cup \left[\tfrac{3}{4},\, 1\right],\, \left[\tfrac{1}{4},\, \tfrac{1}{2}\right) \cup \left[\tfrac{3}{4},\, 1\right],\, [0,\, 1] \right\}. \end{gathered}$$

Now consider the experiment of repeated fair coin flips: $c_t \in \{0,\ 1\}$*. The infinite sequence* $\{c_t\}_{t=0}^{\infty}$ *is a stochastic process that can be modeled with the probability space and associated sequence of* σ*-algebras that we have defined above. Each sequence* $\{c_t\}_{t=0}^{\infty}$ *is an "outcome", represented by a number* $\omega \in \Omega$*.*

For every t *let* $y_t = \{c_j\}_{j=1}^{t}$ *(this will be a* t*-dimensional vector of zeros and ones), and to each possible configuration of* y_t *(there are* 2^t *possible ones), associate a distinct interval in* A_t*. For example, for* $t = 1$ *and* $t = 2$*, let*

$$\begin{aligned} I_1[(0)] &= \left[0,\, \tfrac{1}{2}\right) \\ I_1[(1)] &= \left[\tfrac{1}{2},\, 1\right] \\ I_2[(0,\, 0)] &= \left[0,\, \tfrac{1}{4}\right) \\ I_2[(0,\, 1)] &= \left[\tfrac{1}{4},\, \tfrac{1}{2}\right) \\ I_2[(1,\, 0)] &= \left[\tfrac{1}{2},\, \tfrac{3}{4}\right) \\ I_2[(1,\, 0)] &= \left[\tfrac{3}{4},\, 1\right]. \end{aligned}$$

For $t = 3$*, we will have a three-coordinate vector, and we will have the following restrictions on* I_3*:*

$$\begin{aligned} I_3[(0,\, 0,\, \cdot)] &\subset \left[0,\, \tfrac{1}{4}\right) \\ I_3[(0,\, 1,\, \cdot)] &\subset \left[\tfrac{1}{4},\, \tfrac{1}{2}\right) \\ I_3[(1,\, 0,\, \cdot)] &\subset \left[\tfrac{1}{2},\, \tfrac{3}{4}\right) \\ I_3[(1,\, 1,\, \cdot)] &\subset \left[\tfrac{3}{4},\, 1\right] \end{aligned}$$

and so on for the following t.

Then a number $\omega \in \Omega$ implies a sequence of intervals $\{I_t\}_{t=0}^{\infty}$ that represents, for every t, the "partial" outcome realized that far.

Finally, the stochastic process will be modeled by a function x_t that, for each t and for each $\omega \in \Omega$, associates a real number; such that x_t is $\mathcal{F}_t$-measurable. For example, take $\omega' = .7$ and $\omega'' = .8$, then $I_1[y_1'] = I_1[y_1''] = \left[\frac{1}{2}, 1\right]$ - that is, the first element of the respective sequences c_t', c_t'' is a 1 (say "Heads"). It holds that we must have $x_1(\omega') = x_1(\omega'') \equiv b$.

We are now ready to answer the following question: What is the probability that the first toss in the experiment is "Heads"? Or, in our model, what is the probability that $x_1(\omega) = b$? To answer this question, we look at measure of the set of ω that will produce the value $x_1(\omega) = b$:

$$E = \{\omega : x_1(\omega) = b\} = \left[\tfrac{1}{2}, 1\right] \quad (\in \mathcal{F}_1)$$

The probability of the event $\left[\frac{1}{2}, 1\right]$ is calculated using $P\left(\left[\frac{1}{2}, 1\right]\right) = \lambda\left(\left[\frac{1}{2}, 1\right]\right) = \frac{1}{2}$. That is, the probability that the event $\{c_t\}_{t=1}^{\infty}$ to be drawn produces a Head as its first toss is $\frac{1}{2}$.

Definition 6.17 *Let $B \in \mathcal{F}$. Then the* ***joint probability*** *of the events $(x_{t+1}, ..., x_{t+n}) \in B$ is given by*

$$P_{t+1,...,t+n}(B) = P\left[\omega \in \Omega : [x_{t+1}(\omega), ..., x_{t+n}(\omega)] \in B\right].$$

Definition 6.18 *A stochastic process is* ***stationary*** *if $P_{t+1,...,t+n}(B)$ is independent of t, $\forall t$, $\forall n$, $\forall B$.*

Conceptually, if a stochastic process is stationary, then the joint probability distribution for any $(x_{t+1}, ..., x_{t+n})$ is independent of time.

Given an observed realization of the sequence $\{x_j\}_{j=1}^{\infty}$ in the last s periods $(x_{t-s}, ..., x_t) = (a_{t-s}, ..., a_t)$, the *conditional* probability of the event $(x_{t+1}, ..., x_{t+n}) \in B$ is denoted by

$$P_{t+1,...,t+n}\left[B \,|x_{t-s} = a_{t-s}, ..., x_t = a_t\right].$$

Definition 6.19 *A first order Markov Process is a stochastic process with the property that*

$$P_{t+1,...,t+n}\left[B \,|x_{t-s} = a_{t-s}, ..., x_t = a_t\right] = P_{t+1,...,t+n}\left[B \,|x_t = a_t\right].$$

Definition 6.20 *A stochastic process is weakly stationary (or covariance stationary) if the first two moments of the joint distribution of $(x_{t+1}, ..., x_{t+n})$ are independent of time.*

A usual assumption in macroeconomics is that the exogenous randomness affecting the economy can be modelled as a (weakly) stationary stochastic process. The task then is to look for stochastic processes for the endogenous variables (capital, output, etc.) that are stationary. This stochastic stationarity is the analogue to the steady state in deterministic models.

Example 6.21 *Suppose that productivity is subject to a two-state shock*

$$\begin{aligned} y &= zF(k) \\ z &\in \{z_L,\, z_H\}\,. \end{aligned}$$

Imagine for example that the z_t's are iid, with $\Pr[z_t = z_H] = \frac{1}{2} = \Pr[z_t = z_L]\ \forall t$*. The policy function will now be a function of both the initial capital stock K and the realization of the shock z, i.e. $g(k,\, z) \in \{g(k,\, z_L),\, g(k,\, z_H)\}\ \ \forall K$. We need to find the functions $g(k,\, \cdot)$. Notice that they will determine a stochastic process for capital,i.e. the trajectory of capital in this economy will be subject to a random shock. The Figure 6.4 shows an example of such a trajectory.*

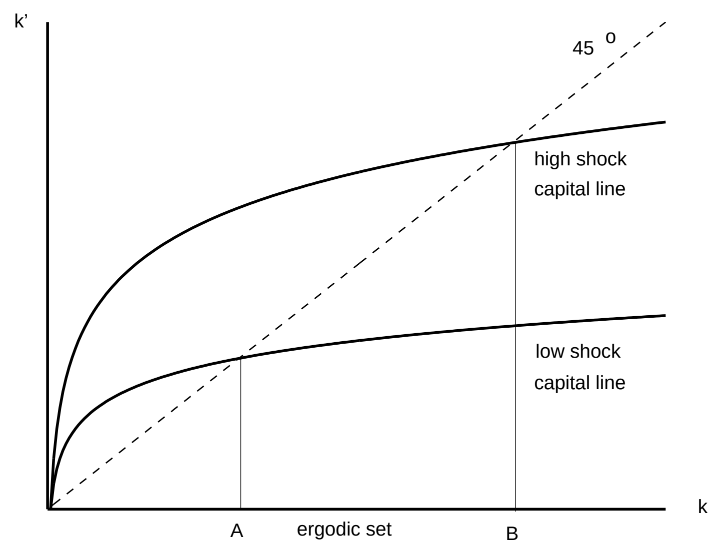

Figure 6.4: Stochastic levels of capital. The interval (A, B) is the *ergodic set*: once the level of capital enters this set, it will not leave it again. The capital stock will follow a stationary stochastic process within the limits of the ergodic set.

Chapter 7

Welfare in macroeconomic models

[INTRO: A QUANTITATIVE AIM. BUT BASIC THEOREMS!]

The infinite-horizon representative-agent (dynastic) model studied above takes a specific view on bequests: bequests are like saving, i.e., purposeful postponements of consumption for later years, whether that later consumption will be for the same person or for other persons in the same "dynasty". An obvious (radical) alternative is the view that people do not value their offspring at all, and that people only save for life-cycle reasons. The present chapter will take that view.

Thus, one motivation for looking at life-cycle savings—and the overlapping-generations, or OG, economy—is as a plausible alternative, on a descriptive level, to the dynastic model. However, it will turn out that this alternative model has a number of quite different features. One has to do with welfare, so a key objective here will be to study the efficiency properties of competitive equilibrium under such setups. In particular, we will demonstrate substantial modifications of the welfare properties of equilibria if "overlapping-generations features" are allowed. Uncertainty will be considered as well, since OG models with uncertainty demand a new discussion of welfare comparisons. Another result is that competitive equilibria, even in the absence of externalities, policy, or nonstandard preferences or endowments, may not be unique. A third feature is that the ownership structure of resources may be important for allocations, even when preferences admit aggregation within each cohort. Finally, the OG economy is useful for studying a number of applied questions, especially those having to do with intertemporal transfer issues from the perspective of the government budget: social security, and the role of government budget deficits.

7.1 Definitions and notation

In what follows, we will introduce some general definitions. By assuming that there is a finite set $\mathcal{H}$ of consumers (and, abusing notation slightly, let H be an index set, such that $H \equiv card(\mathcal{H})$), we can index individuals by a subscript $h = 1, ..., H$. So H agents are born each period t, and they all die in the end of period $t + 1$. Therefore, in each period t the young generation born at t lives together with the "old" people born at $t - 1$.

Let $c_t^h(t+i)$ denote consumption at date $t+i$ of agent h born at t (usually we say "of generation t"), and we have the following:

Definition 7.1 *A **consumption allocation** is a sequence*

$$c = \left\{ \left(c_t^h(t),\, c_t^h(t+1) \right)_{h \in H} \right\}_{t=0}^{\infty} \cup \left(c_{-1}^h(0) \right)_{h \in H}.$$

A consumption allocation defines consumption of agents of all generations from $t = 0$ onwards, including consumption of the initial old, in the economy.

Let $c(t) \equiv \sum_{h \in H} \left[c_t^h(t) + c_{t-1}^h(t) \right]$ denote total consumption at period t, composed of the amount $c_t^h(t)$ consumed by the *young* agents born at t, and the consumption $c_{t-1}^h(t)$ enjoyed by the *old* agents born at $t-1$. Then we have the following:

Example 7.2 (Endowment economy) *In an endowment economy, a consumption allocation is **feasible** if*

$$c(t) \leq Y(t) \quad \forall t.$$

Example 7.3 (Storage economy) *Assume there is "intertemporal production" modelled as a storage technology whereby investing one unit at t yields γ units at $t+1$. In this case, the application of the previous definition reads: a consumption allocation is feasible in this economy if there exists a sequence $\{K(t)\}_{t=0}^{\infty}$ such that*

$$c(t) + K(t+1) \leq Y(t) + K(t)\gamma \quad \forall t,$$

where $Y(t)$ is an endowment process.

Example 7.4 (Neoclassical growth model) *Let $L(t)$ be total labor supply at t, and the neoclassical function $Y(t)$ represent production technology:*

$$Y(t) = F\left[K(t),\, L(t)\right].$$

Capital is accumulated according to the following law of motion:

$$K(t+1) = (1-\delta)\, K(t) + I(t).$$

Then in this case (regardless of whether this is a dynastic or an overlapping generations setup), we have that a consumption allocation is feasible if there exists a sequence $\{I(t)\}_{t=0}^{\infty}$ such that

$$c(t) + I(t) \leq F\left[K(t),\, L(t)\right] \quad \forall t.$$

The definitions introduced so far are of *physical* nature: they refer only to the material possibility to attain a given consumption allocation. We may also want to open judgement on the desirability of a given allocation. Economists have some notions to accommodate this need, and to that end we introduce the following definition:

Definition 7.5 *A feasible consumption allocation c is **efficient** if there is no alternative feasible allocation $\widehat{c}$ such that*

$$\widehat{c}(t) \geq c(t) \quad \forall t, \text{ and}$$
$$\widehat{c}(t) > c(t) \text{ for some } t.$$

An allocation is thus deemed efficient if resources are not wasted; that is, if there is no way of increasing the total amount consumed in some period without decreasing consumption in the remaining periods.

The previous definition, then, provides a tool for judging the "desirability" of an allocation according to the aggregate consumption pattern. The following two definitions allow an extension of economists' ability to assess this desirability to the actual distribution of goods among agents.

Definition 7.6 *A feasible consumption allocation c_A is **Pareto superior** to c_B (or c_A "Pareto dominates" c_B) if*

1. *No agent strictly prefers the consumption path specified by c_B to that specified by c_A:*

$$c_A \succsim_{h,t} c_B \quad \forall h \in H, \forall t.$$

2. *At least one agent strictly prefers the allocation c_A to c_B :*

$$\exists j \in H, \widehat{t} : c_A \succ_{j,\widehat{t}} c_B.$$

Notice that this general notation allows each agent's preferences to be defined on other agents' consumption, as well as on his own. However, in the overlapping-generations model that we will study the agents will be assumed to obtain utility (or disutility) only from their own consumption. Then, condition for Pareto domination may be further specified. Define $c_t^h = \left\{c_t^h(t), c_t^h(t+1)\right\}$ if $t \geq 0$ and $c_t^h = \left\{c_t^h(t+1)\right\}$ otherwise. Pareto domination condition reads:

1. No agent strictly prefers his/her consumption path implied by c_B to that implied by c_A:

$$c_{A_t}^h \succsim_{h,t} c_{B_t}^h \quad \forall h \in H, \forall t.$$

2. At least one agent strictly prefers the allocation c_A to c_B:

$$\exists j \in H, \widehat{t} : c_{A_{\widehat{t}}}^j \succ_{j,\widehat{t}} c_{B_{\widehat{t}}}^j.$$

Whenever c_B is implemented, the existence of c_A implies that a welfare improvement is feasible by modifying the allocation. Notice that a welfare improvement in this context means that it is possible to provide at least one agent (and potentially many of them) with a consumption pattern that he will find preferable to the *status quo*, while the remaining agents will find the new allocation at least as good as the previously prevailing one.

Building on the previous definition, we can introduce one of economists' most usual notions of the most desirable allocation that can be achieved in an economy:

Definition 7.7 *A consumption allocation c is **Pareto optimal** if:*

1. *It is feasible.*

2. *There is no other feasible allocation $\widehat{c} \neq c$ that Pareto dominates c.*

Even though we accommodated the notation to suit the overlapping-generations framework, the previous definitions are also applicable to the dynastic setup. In what follows we will restrict our attention to the overlapping-generations model to study the efficiency and optimality properties of competitive equilibria. You may suspect that the fact that agents' life spans are shorter than the economy's horizon might lead to a different level of capital accumulation than if agents lived forever. In fact, a quite general result is that economies in which generations overlap lead to an overaccumulation of capital. This is a form of (dynamic) inefficiency, since an overaccumulation of capital implies that the same consumption pattern could have been achieved with less capital investment – hence more goods could have been "freed-up" to be consumed.

In what follows, we will extend the concept of competitive equilibrium to the overlapping generations setup. We will start by considering endowment economies, then extend the analysis to production economies, and finally to the neoclassical growth model.

7.2 An endowment economy

We continue to assume that agents of every generation are indexed by the index set H. Let $\omega_t^h(t+i)$ denote the endowment of goods at $t+i$ of agent h born at t. Then the total endowment process is given by

$$Y(t) = \sum_{h \in H} \omega_t^h(t) + \omega_{t-1}^h(t).$$

We will assume throughout that preferences are strongly monotone which means that all inequality constraints on consumption will bind.

7.2.1 Sequential markets

We assume that contracts between agents specifying one-period loans are enforceable, and we let $R(t)$ denote the gross interest rate for loans granted at period t and maturing at $t+1$. Then each agent h born at $t \geq 0$ must solve

$$\max_{c_1,\, c_2} u_t^h\left(c_1,\, c_2\right) \tag{7.1}$$

$$\begin{aligned} \text{s.t.}\quad & c_1 + l \leq \omega_t^h(t), \\ & c_2 \leq \omega_t^h(t+1) + lR(t), \end{aligned}$$

and generation -1 trivially solves

$$\max_{c_{-1}^h(0)} u_{-1}^h\left[c_{-1}^h(0)\right] \tag{7.2}$$

$$\text{s.t. } c_{-1}^h(0) \leq \omega_{-1}^h(0).$$

Unlike the dynastic case, there is no need for a no-Ponzi game restriction. In the dynastic model, agents could keep on building debt forever, unless prevented to do so. But now, they must repay their loans before dying, which happens in finite time[1].

Definition 7.8 *A* ***competitive equilibrium with sequential markets*** *is a consumption allocation c and a sequence $R \equiv \{R(t)\}_{t=0}^{\infty}$ such that*

1. $\left(c_t^h(t),\, c_t^h(t+1)\right)$ *solve generation t's agent h (7.1) problem, and $c_{-1}^h(0)$ solves (7.2) problem.*

2. *Market clearing is satisfied. (Effectively, we need only to require the credit market to be cleared, and Walras' law will do the rest due to feasibility of c):*

$$\sum_{h\in H} l_t^h = 0, \forall t = 0, \ldots, +\infty.$$

In the initial setup of the model the agents were assumed to live for two periods. Because of this, no intergenerational loan can be ever paid back (either a borrower, or a lender is simply not there next period). Therefore, there is no intergenerational borrowing in the endowment economy.

7.2.2 Arrow-Debreu date-0 markets

In this setup we assume that all future generations get together at date $t = -1$ in a futures market and arrange delivery of consumption goods for the periods when they will live[2].

The futures market to be held at $t = -1$ will produce a price sequence $\{p(t)\}_{t=0}^{\infty}$ of future consumption goods. Then each consumer (knowing in advance the date when he will be reborn to enjoy consumption) solves

$$\max_{c_1,\, c_2} u_t^h\left(c_1,\, c_2\right). \tag{7.3}$$

$$\text{s.t. } p(t)c_1 + p(t+1)c_2 \leq p(t)\omega_t^h(t) + p(t+1)\omega_t^h(t+1)$$

[1]Notice that in fact both the no-Ponzi-game and this "pay-before-you-die" restrictions are of an institutional nature, and they play a key role in the existence of an inter-temporal market – the credit market.

[2]You may assume that they all sign their trading contracts at $t = -1$, thereafter to die immediately and be reborn in their respective periods – the institutional framework in this economy allows enforcement of contracts signed in previous lives.

whenever his next life will take place at $t \geq 0$, and the ones to be born at $t = -1$ will solve

$$\max_c u_0^h(c) \tag{7.4}$$
$$\text{s.t. } p(0)c \leq p(0)\omega_{-1}^h(0).$$

Definition 7.9 *A* ***competitive equilibrium with Arrow-Debreu date-0 markets*** *is a consumption allocation c and a sequence $p \equiv \{p(t)\}_{t=0}^{\infty}$ such that*

1. *$\left(c_t^h(t),\, c_t^h(t+1)\right)$ solve generation t's agent h (7.3) problem, and $c_{-1}^h(0)$ solves (7.4) problem.*

2. *Resource feasibility is satisfied (markets clear).*

Claim 7.10 *The definitions of equilibrium with sequential markets and with Arrow-Debreu date-0 trading are equivalent. Moreover, if $(c,\, p)$ is an Arrow-Debreu date-1 trading equilibrium, then $(c,\, R)$ is a sequential markets equilibrium where*

$$R(t) = \frac{p(t)}{p(t+1)}. \tag{7.5}$$

Proof. Recall the sequential markets budget constraint of an agent born at t:

$$c_1 + l = \omega_t^h(t),$$
$$c_2 = \omega_t^h(t+1) + lR(t),$$

where we use the strong monotonicity of preferences to replace the inequalities by equalities. Solving for l and replacing we obtain:

$$c_1 + \frac{c_2}{R(t)} = \omega_t^h(t) + \frac{\omega_t^h(t+1)}{R(t)}.$$

Next recall the Arrow-Debreu date-0 trading budget constraint of the same agent:

$$p(t)c_1 + p(t+1)c_2 = p(t)\omega_t^h(t) + p(t+1)\omega_t^h(t+1).$$

Dividing through by $p(t)$, we get

$$c_1 + \frac{p(t+1)}{p(t)}c_2 = \omega_t^h(t) + \frac{p(t+1)}{p(t)}\omega_t^h(t+1).$$

As can be seen, with the interest rate given by (7.5) the two budget sets are identical. Hence comes the equivalence of the equilibrium allocations.

An identical argument shows that if $(c,\, R)$ is a sequential markets equilibrium, then $(c,\, p)$ is an Arrow-Debreu date-0 trading equilibrium, where prices $p(t)$ are determined by normalizing $p(0) = p_0$ (usual normalization is $p_0 = 1$) and deriving the remaining ones recursively from

$$p(t+1) = \frac{p(t)}{R(t)}.$$

■

Remark 7.11 *The equivalence of the two equilibrium definitions requires that the amount of loans that can be drawn, l, be unrestricted (that is, that agents face no borrowing constraints other than the ability to repay their debts). The reason is that we can switch from*

$$c_1 + l = \omega_t^h(t)$$
$$c_2 = \omega_t^h(t+1) + lR(t)$$

to

$$c_1 + \frac{c_2}{R(t)} = \omega_t^h(t) + \frac{\omega_t^h(t+1)}{R(t)} \tag{7.6}$$

only in the absence of any such restrictions.

Suppose instead that we had the added requirement that $l \geq b$ for some number b such that $b > -\frac{\omega_t^h(t+1)}{R(t)}$. In this case, (7.11) and (7.6) would not be identical any more[3].

7.2.3 Application: endowment economy with one agent per generation

We will assume that $H = 1$ (therefore agents are now in fact indexed only by their birth dates), and that for every generation $t \geq 0$ preferences are represented by the following utility function:

$$u_t(c_y, c_o) = \log c_y + \log c_o.$$

Similarly, the preferences of generation $t = -1$ are represented by utility function

$$u_{-1}(c) = \log c.$$

The endowment processes are given by:

$$\begin{aligned} \omega_t(t) &= \omega_y, \\ \omega_t(t+1) &= \omega_o. \end{aligned}$$

for all t. Trading is sequential, and there are no borrowing constraints other than solvency.

Agent $t \geq 0$ now solves

$$\max_{c_y,\, c_o} \log c_y + \log c_o$$

s.t.

$$c_y + \frac{c_o}{R(t)} = \omega_y + \frac{\omega_o}{R(t)}.$$

We can substitute for c_o to transform the agent's problem into:

$$\max_{c_y} \log c_y + \log\left[\left(\omega_y + \frac{\omega_o}{R(t)} - c_y\right) R(t)\right].$$

[3] If $b = -\frac{\omega_t^h(t+1)}{R(t)}$, then this is just the "pay-before-you-die" restriction - implemented in fact by non-negativity of consumption. Also, if $b < -\frac{\omega_t^h(t+1)}{R(t)}$, then $l \geq b$ would never bind, for the same reason.

Taking first-order conditions yields:

$$\frac{1}{c_y} - \frac{R(t)}{\left(\omega_y + \frac{\omega_o}{R(t)} - c_y\right) R(t)} = 0,$$
$$c_y = \omega_y + \frac{\omega_o}{R(t)} - c_y.$$

Then, from first-order condition and budget constraint we get:

$$\begin{aligned} c_y &= \frac{1}{2}\left(\omega_y + \frac{\omega_o}{R(t)}\right), \\ c_o &= \frac{1}{2}\left(\omega_y R(t) + \omega_o\right). \end{aligned}$$

Market clearing and strong monotonicity of preferences require that the initial old consume exactly their endowment:

$$c_{-1}(0) = \omega_o.$$

Therefore, using the feasibility constraint for period $t = 0$, that reads:

$$c_0(0) + c_{-1}(0) = \omega_y + \omega_o,$$

follows:

$$c_0(0) = \omega_y[4].$$

Repeating the market clearing argument for the remaining t (since $c_0(0) = \omega_y$ will imply $c_0(1) = \omega_o$), we obtain the following equilibrium allocation, $\forall t$:

$$\begin{aligned} c_t(t) &= \omega_y, \\ c_t(t+1) &= \omega_o. \end{aligned}$$

Given this allocation, we solve for the prices $R(t)$ that support it. You may check that these are

$$R(t) = \frac{\omega_o}{\omega_y}.$$

This constant sequence supports the equilibrium where agents do not trade: they just consume their initial endowments.

Let us now use specific numbers to analyze a quantitative example. Let

$$\begin{aligned} \omega_y &= 3, \\ \omega_o &= 1. \end{aligned}$$

[4] Notice that the same result follows from clearing of the loans market at $t = 0$: $l_0 = 0$. This, together with $c_0(0) + l_0 = \omega_y$, implies the same period 0 allocation.

This implies the *gross* interest rate of $R(t) = \frac{1}{3}$. The *net* interest rate is negative: $r(t) \equiv R(t) - 1 = -\frac{2}{3}$.

The natural question, hence, is whether the outcome $R(t) = \frac{1}{3}$ is a) efficient; and b) optimal:

a) Efficiency: Total consumption under the proposed allocation is $c(t) = 4$, which is equal to the total endowment. It is not possible to increase consumption in any period because there is no waste of resources. Therefore, the allocation is *efficient*.

b) Optimality: To check whether the allocation is optimal, consider the following alternative allocation:

$$\begin{aligned}\widehat{c}_{-1}(0) &= 2,\\ \widehat{c}_t(t) &= 2,\\ \widehat{c}_t(t+1) &= 2.\end{aligned}$$

That is, the allocation $\widehat{c}$ is obtained from a chain of intergenerational good transfers that consists of the young in every period giving a unit of their endowment to the old in that period. Notice that for all generations $t \geq 0$, this is just a modification of the timing in their consumption, since total goods consumed throughout their lifetime remain at 4. For the initial old, this is an increase from 1 to 2 units of consumption when old. It is clear, then, that the initial old strictly prefer $\widehat{c}$ to c. We need to check what the remaining generations think about the change. It is clear that since utility is concave (the log function is concave), this even split of the same total amount will yield a higher utility value. In fact,

$$u_t\left(\widehat{c}_t\right) = \log 2 + \log 2 = 2 \cdot \log 2 = \log 4 > \log 3 + \log 1 = \log 3 = u_t\left(c_t\right).$$

Therefore, $\widehat{c}$ Pareto dominates c, which means that c can not be Pareto optimal.

Suppose instead that the endowment process is reversed in the following way:

$$\begin{aligned}\omega_y &= 1,\\ \omega_o &= 3.\end{aligned}$$

There is the same total endowment in the economy each period, but the relative assignments of young and old are reversed. From the formula that we have derived above, this implies

$$R(t) = 3.$$

The “no trade” equilibrium where each agent consumes his own endowment each period is efficient again, since no goods are wasted.

Is it Pareto optimal? This seems a difficult issue to address, since we need to compare the prevailing allocation with all other possible allocations. We already know that an allocation having (2, 2) will be preferred to (1, 3) given the log utility assumption. However, is it possible to start a sequence of intergenerational transfers achieving consumption of (c_y, c_o) from some t (≥ 0) onwards, while keeping the constraints that all generations receive at least $\log 3$ units of utility throughout their lifetime, some generation is strictly better off, and the initial old consume at least 3 units? (If any of these constraints is violated, the allocation thus obtained will not Pareto dominate the "no trade" allocation.) We will provide an answer to this question.

We will first restrict attention to alternative *stationary* allocations. Let us introduce a more formal definition of this term.

Definition 7.12 (Stationary allocation) *A feasible allocation c is called **stationary** if* $\forall t$:

$$\begin{aligned} c_t(t) &= c_y, \\ c_t(t+1) &= c_o. \end{aligned}$$

With this definition at hand, we can pose the question of whether there is any *stationary* allocation that Pareto dominates (2, 2). Figure 7.1 shows the resource constraint of the economy, plotted together with the utility level curve corresponding to the allocation (2, 2):

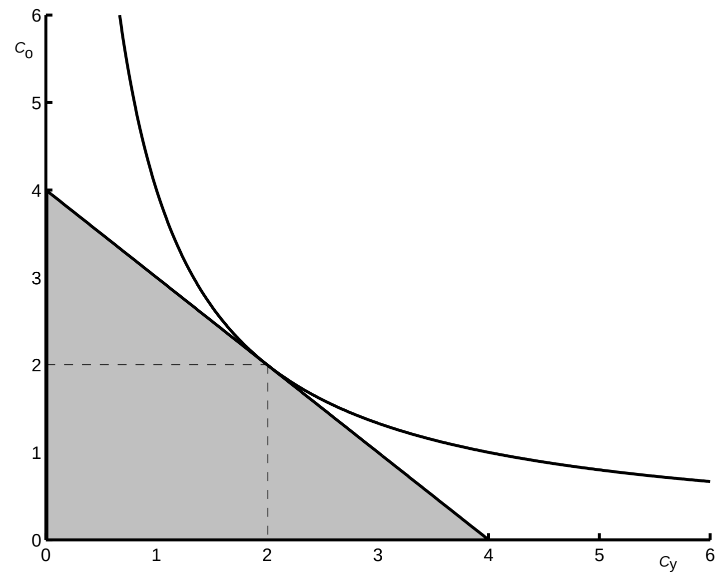

Figure 7.1: Pareto optimality of (2, 2) allocation

The shaded area is the feasible set, its frontier given by the line $c_y + c_o = 4$. It is clear from the tangency at (2, 2) that it is not possible to find an alternative allocation that Pareto dominates this one. However, what happens if we widen our admissible range of allocations and think about non-stationary ones? Could there be a non-stationary allocation dominating (2, 2)?

In order to implement such a non-stationary allocation, a chain of inter-generational transfers would require a transfer from young to old at some arbitrary point in time t. These agents giving away endowment units in their youth would have to be compensated when old. The question is how many units of goods would be required for this compensation.

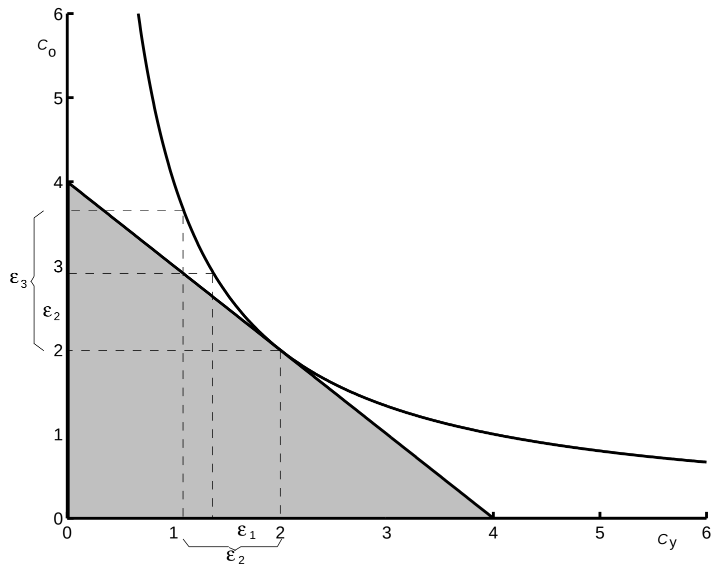

Figure 7.2: Impossibility of Pareto improvement over (2, 2) allocation

Figure 7.2 illustrates that, given an initial transfer ε_1 from young to old at t, the transfer ε_2 required to compensate generation t must be larger than ε_1, given the concave utility assumption. This in turn will command a still larger ε_3, and so on. Is the sequence $\{\varepsilon_t\}_{t=0}^{\infty}$ thus formed feasible?

An intuitive answer can be seen in the chart: no such transfer scheme is feasible in the long run with stationary endowment process. Therefore, for this type of preferences the stationary allocation (2, 2) is the Pareto optimal allocation. Any proposed non-stationary allocation that Pareto dominates (2, 2) becomes unfeasible at some point in time.

Somewhat more formally, let us try to use the First Welfare Theorem to prove Pareto optimality. Notice that our model satisfies the following key assumption:

- Preferences exhibit local non-satiation (since u is strictly increasing).

Proof (Pareto optimality of competitive equilibrium). Let an economy's population be indexed by a countable set I (possibly infinite), and consider a competitive equilibrium allocation x that assigns x_i to each agent i (x_i might be multi-dimensional).

If x is not Pareto optimal, then there exists $\widehat{x}$ that Pareto dominates x, that is, a feasible allocation that satisfies:

$$\forall i \in I : \widehat{x}_i \succeq_i x_i,$$

$$\exists j \in I : \widehat{x}_j \succ_j x_j.$$

Then we can use local non-satiation to show that

$$\begin{aligned} p\widehat{x}_i &\geq px_i, \\ p\widehat{x}_j &> px_j \end{aligned}$$

must hold.

Summing up over all agents, we get

$$\begin{aligned} \sum_{i \in I} p\widehat{x}_i &> \sum_{i \in I} px_i, \\ p \sum_{i \in I} \widehat{x}_i &> p \sum_{i \in I} x_i. \end{aligned}$$

The last inequality violates the market clearing condition, since the market value of goods (with local non-satiation) must be equal to the market value of endowments in an equilibrium. ■

This proof is quite general. In the specific case of infinite-horizon models, overlapping generations, we have two peculiarities: p and x are infinite-dimensional vectors. Do they cause problems in the proof? As long as the px products and the summations are finite, no. In fact, in any competitive equilibrium of the dynastic model of previous chapters, these products are by definition finite, since they define consumers' budget sets, and the "maximization" part of their problems would not be met were budgets infinite.

In a two-period-life overlapping-generations economy, individuals' budgets are finite as well: the x vector contains just two (finite) elements, with the remaining entries set at zero. However, a new complication arises: there is an infinite set of consumers. Therefore, the series $\sum\limits_{i \in I} p\widehat{x}_i$ and $\sum\limits_{i \in I} px_i$ might take on an infinite value, in which case the last comparison in the proof might not hold. We need to specify further conditions to ensure that the first welfare theorem will hold, even with the "correct" assumptions on preferences. Thus, in a competitive equilibrium of an OG economy where the states sums are well defined, the above proof can be used. But there are other cases as well, and for these cases, more analysis is needed.

To this effect, let us assume that the following conditions are met by the economy:

1. Regularity conditions on utility and endowments.

2. Restrictions on the curvature of the utility function – that has to be "somewhat" curved, but not too much. An example of curvature measure is (one over) the elasticity of intertemporal substitution:

$$-\frac{f''(x)x}{f'(x)}.$$[5]

[5]This ratio is also called the coefficient of relative risk aversion whenever the environment involves uncertainty. In the expected utility framework the same ratio measures two aspects of preferences: intertemporal comparison, and degree of aversion to stochastic variability of consumption.

3. Other technical details that you may find in Balasko and Shell (1980).

Then we have the following:

Theorem 7.13 (Balasko and Shell, Journal of Economic Theory, 1980) *A competitive equilibrium in an endowment economy populated by overlapping generations of agents is Pareto optimal if and only if*

$$\sum_{t=0}^{\infty} \frac{1}{p(t)} = \infty,$$

where $p(t)$ denote Arrow-Debreu prices for goods delivered at time t.

Recall our example. The allocation (2, 2) implied $R(t) = 1$, and from the equivalence of sequential and Arrow-Debreu date-0 trading equilibria, we have that

$$p(t+1) = \frac{p(t)}{R(t)},$$

which implies

$$\sum_{t=0}^{\infty} \frac{1}{p(t)} = \sum_{t=1}^{\infty} \frac{1}{p(0)} = \infty.$$

In the case of (3, 1), we have

$$p(t) = 3^t \cdot p(0).$$

Then

$$\sum_{t=0}^{\infty} \frac{1}{p(t)} = \sum_{t=0}^{\infty} \frac{3^{-t}}{p(0)} = \frac{1}{p(0)} \sum_{t=0}^{\infty} 3^{-t} = \frac{1}{2 \cdot p(0)} < \infty.$$

And finally for (1, 3),

$$\sum_{t=0}^{\infty} \frac{1}{p(t)} = \sum_{t=0}^{\infty} \frac{3^t}{p(0)} = \infty.$$

Therefore, by applying the theorem we conclude that (2, 2) and (1, 3) are Pareto optimal allocations, whereas (3, 1) can be improved upon, which is the same conclusion we had reached before.

So, what if the economy in question can be represented as (3, 1) type of situation? How can a Pareto improvement be implemented? Should the government step in, and if so, how?

A possible answer to this question is a "pay-as-you-go" type of social security system that is used in many economies worldwide. But a distinct drawback of such a solution is the forced nature of payments, when social security becomes "social coercion". Is there any way to implement Pareto superior allocation with the help of the market?

One of the solutions would be to endow the initial old with (intrinsically useless) pieces of paper called "money". Intuitively, if the initial old can make the young in period $t = 0$ believe that at time $t = 1$ the next young will be willing to trade valuable goods for these pieces of paper, a Pareto improvement can be achieved relying solely on the market forces. We will examine this issue in the following section in greater detail.

7.3 Economies with intertemporal assets

In the previous section, we have looked at overlapping-generations economies in which only consumption goods are traded. A young agent selling part of his endowment to an old one obviously needs something which serves the purpose of a storage of value, so that the proceeds from the sale performed at time t can be used to purchase goods at $t+1$. A unit of account is therefore implicit in the framework of the previous section, which is obvious from the moment that such thing as "prices" are mentioned. However, notice that such units of account are not money, they exist only for convenience of quoting relative prices for goods in different periods.

We will now introduce intertemporal assets into the economy. We will consider in turn fiat money and real assets.

7.3.1 Economies with fiat money

In this section we introduce "fiat" money to the economy. To this end, any paper with a number printed on it will fulfill the need of value storage, provided that everybody agrees on which are the valid papers, and no forgery occurs. We have assumed away these details: agents are honest.

As before, consider an overlapping-generations economy with agents who live for two periods, one agent per generation. An endowment process is given by:

$$(\omega_t(t), \omega_t(t+1)) = (\omega_y, \omega_o), \quad \forall t.$$

The preferences will once again be assumed to be logarithmic:

$$u_t(c_y, c_o) = \log c_y + \log c_o, \ \forall t.$$

In contrast to the previous setup, let the initial old be endowed with M units of fiat currency. A natural question to address is whether money can have value in this economy.

A bit of notation: let p_{mt} denote a value of a unit of money at time t in terms of consumption goods at time t. Also let $p_t \equiv \dfrac{1}{p_{mt}}$ be "price level" at time t, that is, the price of a unit of consumption goods at time t in terms of money. Notice the difference between p_t in this model and Arrow-Debreu date-0 prices denoted $p(t)$.

Assume for the moment that $p_t < \infty$. Then, the maximization problem of generation t agent is:

$$\begin{aligned} &\max_{c_y, c_o, M'} \ \log c_y + \log c_o \\ &\text{s.t.} \quad c_y + \frac{M'}{p_t} = \omega_y, \\ &\qquad\quad c_o = \omega_o + \frac{M'}{p_{t+1}}, \\ &\qquad\quad M' \geq 0. \end{aligned} \tag{7.7}$$

And the agent of generation -1 trivially solves:

$$\max_{c_{-1}(0)} \log c_{-1}(0)$$

$$\text{s.t.} \quad c_{-1}(0) = \omega_o + \frac{M'}{p_0}.$$

The meaning of the last constraint in (7.7) is that agents cannot issue money, or, alternatively, sell it short. Combining the constraints from (7.7), the consolidated budget constraint of an agent born at period t is:

$$c_y + \frac{c_o}{\frac{p_t}{p_{t+1}}} = \omega_y + \frac{\omega_o}{\frac{p_t}{p_{t+1}}},$$
$$\omega_y - c_y \geq 0.$$

The budget set under these constraints is presented in Figure 7.3. As can be seen, the real return on money is $\frac{p_t}{p_{t+1}} \equiv \frac{1}{1+\pi_{t+1}}$. Here π_{t+1} denotes the inflation rate. From first-order Taylor approximation it follows that net real return on one dollar invested in money is $\simeq -\pi_{t+1}$ (for small values of π_{t+1}).

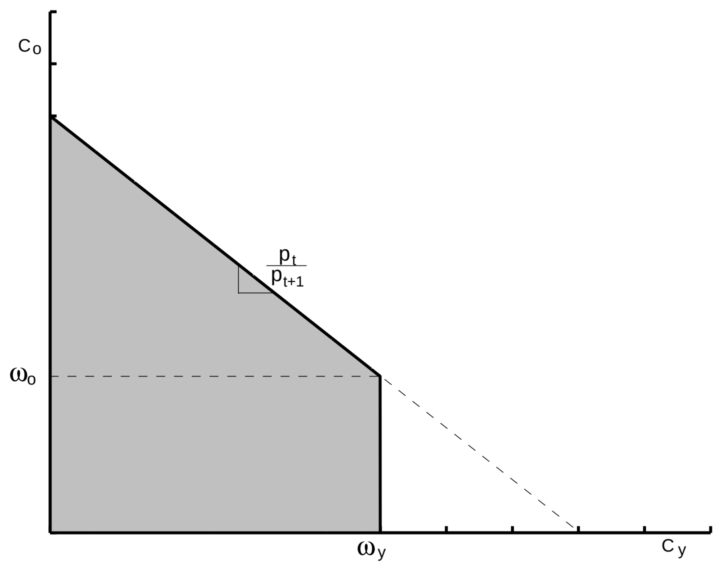

Figure 7.3: Budget set in the economy with fiat money

Momentarily ignore $\omega_y - c_y \geq 0$. Then the solution to (7.7) is:

$$\begin{aligned} c_y &= \frac{1}{2}\left(\omega_y + \omega_o \frac{p_{t+1}}{p_t}\right), \\ c_o &= \frac{1}{2}\left(\omega_y + \omega_o \frac{p_{t+1}}{p_t}\right)\frac{p_t}{p_{t+1}}. \end{aligned}$$

Having found c_y, we can recover the real demand for money of the young at t:

$$\frac{M_{t+1}}{p_t} = \omega_y - c_y = \frac{1}{2}\omega_y - \frac{1}{2}\omega_o\frac{p_{t+1}}{p_t}.$$

Imposing market clearing condition on the money market,

$$M_{t+1} = M \ \forall\ t,$$

we can recover the law of motion for prices in this economy:

$$\frac{M}{p_t} = \omega_y - c_y = \frac{1}{2}\omega_y - \frac{1}{2}\omega_o\frac{p_{t+1}}{p_t} \Rightarrow$$

$$p_{t+1} = p_t\frac{\omega_y}{\omega_o} - \frac{2M}{\omega_o}.$$

Consider the following three cases:

- $\dfrac{\omega_y}{\omega_o} > 1;$
- $\dfrac{\omega_y}{\omega_o} = 1;$
- $\dfrac{\omega_y}{\omega_o} < 1.$

The solution to this first-order difference equation is presented graphically on the Figure 7.4.

As can be seen, the only case consistent with positive and finite values of p_t is the first one, when $\omega_y > \omega_o$.

The following solutions can be identified:

1. If $\omega_y > \omega_o$ we can observe the following: there exists a solution $p_t = \bar{p} > 0$. So, money can have real value!
   (a) Money can "overcome suboptimality" when $\omega_y > \omega_o$ and consumption level is constant ($c_y = c_o = \frac{\omega_y+\omega_o}{2}$), since $\frac{p_t}{p_{t+1}} = 1$ implies that $MRS = 1$, and the resulting allocation is Pareto optimal by Balasko-Shell criterion.
   (b) There is no equilibrium with $p_0 < \bar{p}$, which means that one unit of money at $t = 0$ has value at most $\frac{1}{\bar{p}}$.
   (c) If $p_0 > \bar{p}$, there is an equilibrium, which is the solution to

   $$p_{t+1} = \frac{\omega_y}{\omega_o}p_t - \frac{2M}{\omega_o},$$

   with p_0 given. In this equilibrium, $p_t \to \infty$ ($p_{mt} \to 0$), and $\frac{p_{t+1}}{p_t}$ increases monotonically to $\frac{\omega_y}{\omega_o}$. This is an equilibrium with hyperinflation. Money loses value in the limit.

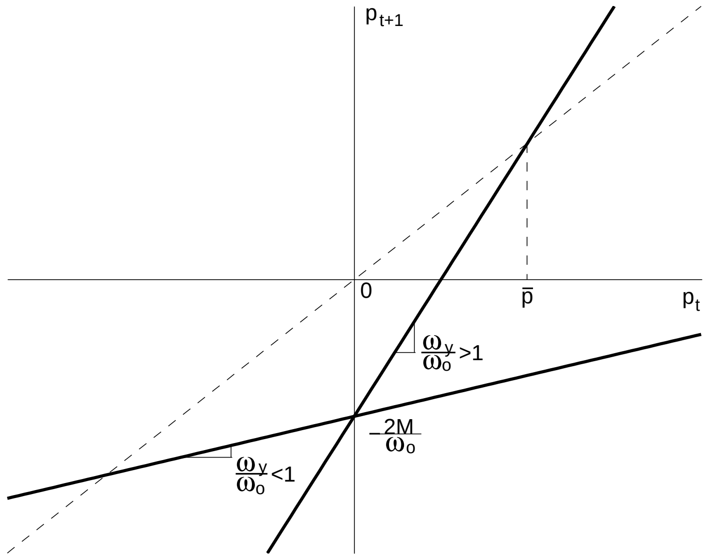

Figure 7.4: Dynamics of price level

(d) $p_{m0} = 0$ ("$p_t = \infty$") is also an equilibrium.

So, there is a continuum of equilibria. The fact that money has value may be seen as a "rational bubble": what people are willing to "pay" for money today depends on what they expect others will "pay" for it tomorrow. The role of money here is to mitigate the suboptimality present in the economy. It is the suboptimality that gives money positive value.

If we add borrowing and lending opportunities, we get from the no-arbitrage condition and market clearing in loans that:

$$R_t = \frac{p_t}{p_{t+1}},\ l_t = 0,\ \forall t.$$

So, real interest rate is non-positive, and (real) money holdings are still present.

2. If $\omega_y \leq \omega_o$ there is no equilibrium with $p_t < \infty$. (However, autarky, $p_t = \infty$, is still an equilibrium.)

Money, in this model, is a store of value. It helps overcome a basic friction in the overlapping-generations model. As we shall see, its role can be filled by other forms of assets as well. This is an important reason why this model of money, though valuable because it is the first model that lets us assign a positive value to an intrinsically useless object (which money is, since it is not backed by gold or other objects in any modern economies), has not survived as a core model of money. A second reason is that, if one were to introduce other assets, such as a "bond", money and bonds would serve similar roles and they would need to carry the same equilibrium return. In reality, however, money has zero (nominal) return, and in this sense bonds dominate money. Thus, this model does not capture another important property of real-world money: its value as a medium of exchange (and its being superior to—more "liquid" than) other assets in this regard.

7.3.2 Economies with real assets

In this subsection we will consider the assets that are real claims, rather than fiat money. That is, they will be actual rights to receive goods in the following periods. Two different kinds of assets are of interest:

- A tree that produces a given fruit yield (dividend) each period.

- Capital, that can be used to produce goods with a given technology.

7.3.3 A tree economy

We assume that the economy is populated by one agent per generation, and that each agent lives for two periods. Preferences are represented by a logarithmic utility function as in previous examples:

$$u_t\left(c_y^t,\, c_o^t\right) = \log c_y^t + \log c_o^t.$$

Agents are endowed with $(\omega_y,\, \omega_o)$ consumption units (fruits) when young and old, respectively, and there is also a tree that produces a fruit yield of d units each period. Therefore total resources in the economy each period are given by:

$$Y(t) = \omega_y + \omega_o + d.$$

Ownership of a given share in the tree gives the right to collect such share out of the yearly fruit produce. Trading of property rights on the tree is enforceable, so any agent that finds himself owning any part of the tree when old will be able to sell it to the young in exchange for consumption goods. The initial old owns 100% of the tree.

Let a_{t+1} denote the share of the tree purchased by the young generation at t, and p_t denotes the price of the tree at t. It is clear that asset market clearing requires $a_{t+1} = 1$ for all t. Generation t consumer solves:

$$\max_{c_y^t,\, c_o^t} \log c_y^t + \log c_o^t$$

$$\text{s.t.} \quad \begin{aligned} & p_t a_{t+1} + c_y^t = \omega_y, \\ & c_o^t = \omega_o + a_{t+1}(p_{t+1} + d). \end{aligned}$$

Notice that the returns on savings are given by

$$\frac{p_{t+1} + d}{p_t}.$$

The first order conditions yield

$$c_y^t = \frac{1}{2}\left(\omega_y + \frac{p_t}{p_{t+1}+d}\omega_o\right),$$

which implies that generation t's savings satisfy:

$$p_t a_{t+1} = \frac{1}{2}\left(\omega_y - \frac{p_t}{p_{t+1}+d}\omega_o\right).$$

Imposing the market clearing condition and rearranging we get the law of motion for prices:

$$p_{t+1} = \frac{\omega_o}{\frac{\omega_y}{p_t} - 2} - d.$$

This is a first order (non-linear) difference equation in p_t. Figure 7.5 shows that it has two fixed points, a stable negative one and an unstable positive one.

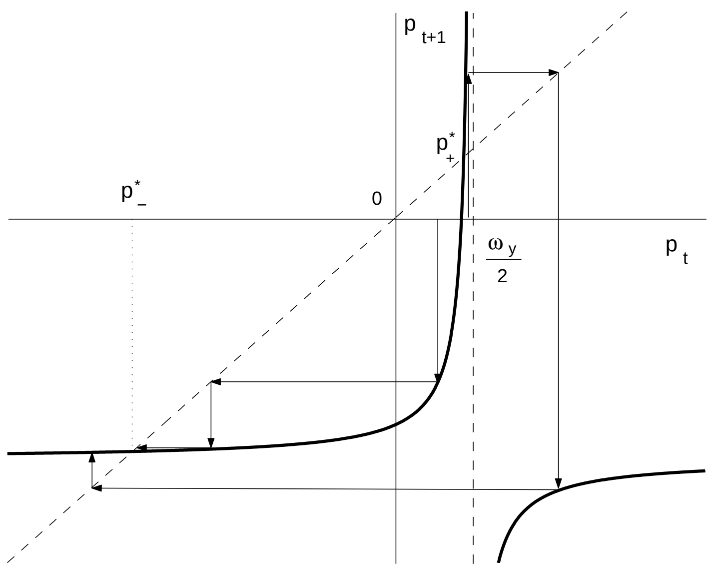

Figure 7.5: Fixed points for price of the tree

What is the equilibrium $\{p_t\}_{t=1}^{\infty}$ sequence? It must be a constant sequence since any deviation from the positive fixed point leads directly into the negative one or creates a "bubble" that eventually collapses due to infeasibility. So, $p_t = p^* \ \forall t$, where p^* is the positive solution to

$$p^* = \frac{\omega_o}{\frac{\omega_y}{p^*} - 2} - d.^6$$

[6] Notice that for the case $d = 0$ we are back in fiat money economy, and the constant positive value of money is once again $p_{mt} = \frac{1}{\bar{p}} = \frac{\omega_y - \omega_o}{2}$ for $M = 1$.

Is this competitive equilibrium Pareto optimal? We can answer this question by checking whether the Balasko-Shell criterion is satisfied. First notice that if we multiply $\frac{1}{p(t)}$ by $\frac{p(t-1)p(t-2)...p(1)p(0)}{p(t-1)p(t-2)...p(1)p(0)}$ we can write:

$$\frac{1}{p(t)} = \frac{p(t-1)p(t-2)...p(1)p(0)}{p(t)p(t-1)...p(1)p(0)} \equiv \prod_{s=0}^{t-1} R_{s,\,s+1},$$

where $p(0) \equiv 1$, and $R_{s,\,s+1}$ denotes the interest rate between periods s and $s+1$:

$$R_{s,\,s+1} \equiv \frac{p(s)}{p(s+1)}.$$

But we already know that the return on savings is given by:

$$\frac{p_{t+1} + d}{p_t}.$$

Therefore, the interest rate for each period, using equilibrium prices, is

$$R_{s,\,s+1} = \frac{p^* + d}{p^*}.$$

Replacing for $\frac{1}{p(t)}$, we get that:

$$\sum_{t=0}^{\infty} \frac{1}{p(t)} = p(0) \sum_{t=0}^{\infty} \left(1 + \frac{d}{p^*}\right)^t.$$

The limit of this series is infinity for any $d \geq 0$. The Balasko-Shell criterion is met; hence, the competitive equilibrium allocation supported by these prices is Pareto optimal.

Finally, notice that the optimality of the result was proven regardless of the actual endowment process; therefore, it generalizes for any such process.

Now consider two cases of economies with production: a simple model with CRS technology that uses only capital, and a more complicated neoclassical growth model.

7.3.4 Storage economy

We will assume the simplest form of production, namely constant marginal returns on capital. Such a technology, represented by a linear function of capital, is what we have called "storage" technology whenever no labor inputs are needed in the production process. Let the yield obtained from storing one unit be equal to one. That is, keeping goods for future consumption involves no physical depreciation, nor does it increase the physical worth of the stored goods.

Let the marginal rates of substitution between consumption when old and when young be captured by a logarithmic function, as before, and assume that the endowment process is $(\omega_y,\, \omega_o) = (3,\, 1)$. Generation t's problem is therefore:

$$\max_{c_y^t,\, c_o^t} \log c_y^t + \log c_o^t$$

$$\text{s.t.} \quad s_t + c_y^t = \omega_y, \\ c_o^t = s_t + \omega_o.$$

The first order conditions yield

$$c_y^t = \frac{1}{2}\left(\omega_y + \frac{\omega_o}{R_t}\right).$$

The return on storage is one, $R_t = 1$. So, using the values assumed for the endowment process, this collapses to

$$\begin{aligned} c_y^t &= 2, \\ c_o^t &= 2, \\ s_t &= 1. \end{aligned}$$

Notice that the allocation corresponds to what we have found to be the Pareto optimal allocation before: (2, 2) is consumed by every agent. In the previous case where no real intertemporal assets existed in the economy, such an allocation was achieved by a chain of intergenerational transfers (enforced, if you like, by the exchange in each period of those pieces of paper dubbed fiat money). Now, however, agent buries his "potato" when young, and consumes it when old.

Is the current allocation Pareto optimal? The answer is clearly no, since, to achieve the consumption pattern (2, 2), the potato must always be buried on the ground. The people who are born at $t = 0$ set aside one unit of their endowment to consume when old, and thereafter all their descendance mimic this behavior, for a resulting allocation

$$c = (1) \cup \{(2,\, 2)\}_{t=0}^{\infty}.$$

However, the following improvement could be implemented. Suppose that instead of storing one, the first generation ($t = 0$) consumed its three units when young. In the following period the new young would give them their own spare unit, instead of storing it, thereafter to continue this chain of intergenerational transfers through infinity and beyond. The resulting allocation would be:

$$\widehat{c} = (1) \cup (3,\, 2) \cup \{(2,\, 2)\}_{t=1}^{\infty},$$

a Pareto improvement on c.

In fact, $\widehat{c}$ is not only a Pareto improvement on c, but simply the same allocation c plus one additional consumption unit enjoyed by generation 0. Since the total endowment of goods is the same, this must mean that one unit was being wasted under allocation c.

This problem is called "overaccumulation of capital". The equilibrium outcome is (dynamically) inefficient.

7.3.5 Neoclassical growth model

The production technology is now modelled by a neoclassical production function. Capital is owned by the old, who put it to production and then sell it to the young each period.

Agents have a labor endowment of ω_y when young and ω_o when old. Assuming that leisure is not valued, generation t's utility maximization problem is:

$$\max_{c_y^t,\, c_o^t} u_t\left(c_y^t,\, c_o^t\right)$$

$$\begin{aligned} \text{s.t.} \quad & c_y^t + s_t = \omega_y w_t, \\ & c_o = s_t r_{t+1} + \omega_o w_{t+1}. \end{aligned}$$

If the utility function is strictly quasiconcave, the savings correspondence that solves this problem is single-valued:

$$s_t = h\left[w_t,\, r_{t+1},\, w_{t+1}\right].$$

The asset market clearing condition is:

$$s_t = K_{t+1}.$$

We require the young at t to save enough to purchase next period's capital stock, which is measured in terms of consumption goods (the price of capital in terms of consumption goods is 1).

The firm operates production technology that is represented by the function $F(K,\, n)$. Market clearing condition for labor is

$$n_t = \omega_y + \omega_o.$$

From the firm's first order conditions of maximization, we have that factor remunerations are determined by

$$\begin{aligned} r_t &= F_1\left(K_t,\, \omega_y + \omega_o\right), \\ w_t &= F_2\left(K_t,\, \omega_y + \omega_o\right). \end{aligned}$$

If we assume that the technology exhibits constant returns to scale, we may write

$$F(K,\, n) = nf\left(\frac{K}{n}\right),$$

where $f\left(\frac{K}{n}\right) \equiv F\left(\frac{K}{n},\, 1\right)$. Replacing in the expressions for factor prices,

$$\begin{aligned} r_t &= f'\left(\frac{K_t}{\omega_y + \omega_o}\right), \\ w_t &= f\left(\frac{K_t}{\omega_y + \omega_o}\right) - \frac{K_t}{\omega_y + \omega_o} f'\left(\frac{K_t}{\omega_y + \omega_o}\right). \end{aligned}$$

Let $k_t \equiv \frac{K_t}{\omega_y + \omega_o}$ denote the capital/labor ratio. If we normalize $\omega_y + \omega_o = 1$, we have that $K_t = k_t$. Then

$$\begin{aligned} r_t &= f'(k_t), \\ w_t &= f(k_t) - k_t f'(k_t). \end{aligned}$$

Substituting in the savings function, and imposing asset market equilibrium,

$$k_{t+1} = h\left[f(k_t) - k_t f'(k_t),\ f'(k_t),\ f(k_{t+1}) - k_{t+1} f'(k_{t+1})\right].$$

We have obtained a *first* order difference equation. Recall that the dynastic model lead to a *second* order equation instead. However, proving convergence to a steady state is usually more difficult in the overlapping generations setup. Recall that the steady state condition with the dynastic scheme was of the form

$$\beta f'(k^*) = 1.$$

In this case, steady state requires that

$$k^* = h\left[f(k^*) - k^* f'(k^*),\ f'(k^*),\ f(k^*) - k^* f'(k^*)\right].$$

7.4 Dynamic efficiency in models with multiple agents

We have analyzed the welfare properties of consumption allocations arising from a multiple agent environment under the form of a population consisting of overlapping generations of individuals. The purpose of this section is to generalize the study of the dynamic efficiency of an economy to a wider range of modelling assumptions. In particular, we will present a theorem valid for any form of one-sector growth model.

We assume that the technology is represented by a neoclassical production function that satisfies the following properties:

- $f(0) = 0$,
- $f'(\cdot) > 0$,
- $f''(\cdot) < 0$,
- $f \in C^2$ (C^2 denotes the space of twice continuously differentiable functions),
- $\lim\limits_{x \to 0} f'(x) = \infty$,
- $\lim\limits_{x \to \infty} f'(x) = 0$.

Notice that since we define $f(x) \equiv F(x, 1) + (1 - \delta)x$, the last assumption is not consistent with the case of $\delta < 1$. This assumption is implicit in what follows. Then we can show the following:

Theorem 7.14 *A steady state k^* is efficient if and only if $R^* \equiv f'(k^*) \geq 1$.*

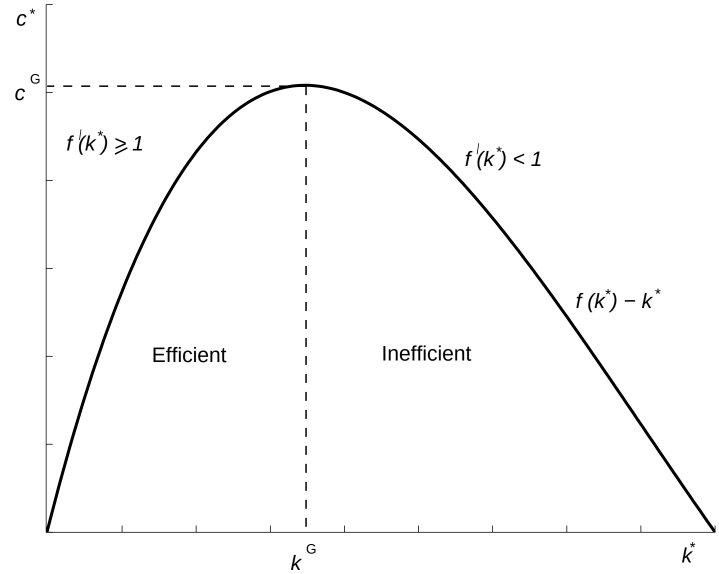

Figure 7.6: Efficiency of the steady state

Intuitively, the steady state consumption is $c^* = f(k^*) - k^*$. Figure 7.6 shows the attainable levels of steady state capital stock and consumption (k^*, c^*), given the assumptions on f. The $\left(k^G, c^G\right)$ locus corresponds to the "golden rule" level of steady state capital and consumption, that maximize c^G.

Proof.

(i) <u>$R^* < 1$: k^* is inefficient.</u>

Assume that k^* is such that $f'(k^*) < 1$. Let c^* denote the corresponding level of steady state consumption, let $c_0 = c^*$. Now consider a change in the consumption path, whereby k_1 is set to $k_1 = k^* - \varepsilon$ instead of $k_1 = k^*$. Notice this implies an increase in c_0. Let $k_t = k_1 \; \forall t \geq 1$. We have that

$$\begin{aligned} c_1 - c^* &= f(k_1) - k_1 - f(k^*) + k^* \\ &\equiv f(k^* - \varepsilon) - (k^* - \varepsilon) - f(k^*) + k^*. \end{aligned}$$

Notice that strict concavity of f implies that

$$f(k^*) < f(k^* - \varepsilon) + [k^* - (k^* - \varepsilon)]\, f'(k^* - \varepsilon)$$

for $\varepsilon \in \left(0, \, k^* - k^G\right)$, and we have that $f'(k^* - \varepsilon) < 1$. Therefore,

$$f(k^*) < f(k^* - \varepsilon) + k^* - (k^* - \varepsilon)\,.$$

This implies that

$$c_1 - c^* > 0,$$

which shows that a permanent increase in consumption is feasible.

(ii) $R^* \geq 1$: k^* is efficient.

Suppose not, then we could decrease the capital stock at some point in time and achieve a permanent increase in consumption (or at least increase consumption at some date without decreasing consumption in the future). Let the initial situation be a steady state level of capital $k_0 = k^*$ such that $f'(k^*) \geq 1$. Let the initial c_0 be the corresponding steady state consumption: $c_0 = c^* = f(k^*) - k^*$. Since we suppose that k^* is inefficient, consider a decrease of capital accumulation at time 0: $k_1 = k^* - \varepsilon_1$, thereby increasing c_0. We need to maintain the previous consumption profile c^* for all $t \geq 1$: $c_t \geq c^*$. This requires that

$$c_1 = f(k_1) - k_2 \geq f(k^*) - k^* = c^*,$$

$$k_2 \leq f(k_1) - f(k^*) + k^*,$$

$$\underbrace{k_2 - k^*}_{\varepsilon_2} \leq f(k_1) - f(k^*).$$

Concavity of f implies that

$$f(k_1) - f(k^*) < f'(k^*)\underbrace{[k_1 - k^*]}_{-\varepsilon_1}.$$

Notice that $\varepsilon_2 \equiv k_2 - k^* < 0$. Therefore, since $f'(k^*) \geq 1$ by assumption, we have that

$$|\varepsilon_2| > |\varepsilon_1|.$$

The size of the decrease in capital accumulation is increasing. By induction, $\{\varepsilon_t\}_{t=0}^{\infty}$ is a decreasing sequence (of negative terms). Since it is bounded below by $-k^*$, we know from real analysis that it must have a limit point $\varepsilon_\infty \in [-k^*, 0)$. Consequently, the consumption sequence converges as well:

$$c_\infty = f(k^* - \varepsilon_\infty) - (k^* - \varepsilon_\infty).$$

It is straightforward to show, using concavity of f, that

$$c_\infty < c^*.$$

Then the initial increase in consumption is not feasible if the restriction is to maintain at least c^* as the consumption level for all the remaining periods of time.

■

We now generalize the theorem, dropping the assumption that the economy is in steady state.

Theorem 7.15 (Dynamic efficiency with possibly non-stationary allocations) *Let both* $\{k_t\}_{t=0}^{\infty}$ *and the associated sequence* $\{R_t(k_t) \equiv f_t'(k_t)\}_{t=0}^{\infty}$ *be uniformly bounded above and below away from zero. Let* $0 < a \leq -f_t''(k_t) \leq M < \infty \quad \forall t, \forall k_t$. *Then* $\{k_t\}_{t=0}^{\infty}$ *is efficient if and only if*

$$\sum_{t=0}^{\infty}\left[\prod_{s=1}^{t} R_s(k_s)\right] = \infty.$$

Recall that

$$\sum_{t=0}^{\infty}\left[\prod_{s=1}^{t} R_s(k_s)\right] = \sum_{t=0}^{\infty}\frac{1}{p_t}.$$

The Balasko-Shell criterion discussed when studying overlapping generations is then a special case of the theorem just presented.

7.5 The Second Welfare Theorem in dynastic settings

From our discussion so far, we can draw the following summary conclusions on the applicability of the first and second welfare theorems to the dynamic economy model.

First Welfare Theorem

1. *Overlapping generations*: Competitive equilibrium is not always Pareto optimal. Sometimes it is not even efficient.

2. *Dynastic model*: Only local non-satiation of preferences and standard assumption $\beta < 1$ are required for competitive equilibrium to be Pareto optimal.

Second Welfare Theorem

1. *Overlapping generations*: In general, there is no applicability of the Second Welfare Theorem.

2. *Dynastic model*: Only convexity assumptions are required for any Pareto optimal allocation to be implementable as a competitive equilibrium.

Therefore with the adequate assumptions on preferences and on the production technology, the dynastic model yields an equivalence between competitive equilibrium and Pareto optimal allocations. Of course, the restrictions placed on the economy for the Second Welfare Theorem to apply are much stronger than those required for the First one to hold. Local non-satiation is almost not an assumption in economics, but virtually the defining characteristic of our object of study (recall that phrase talking about scarce resources, etcetera).

In what follows, we will study the Second Welfare Theorem in the dynastic model. To that effect, we first study a 1-agent economy, and after that a 2-agents one.

7.5.1 The second welfare theorem in a 1-agent economy

We assume that the consumer's preferences over infinite consumption sequences and leisure are represented by a utility function with the following form:

$$U\left[\{c_t,\, l_t\}_{t=0}^{\infty}\right] = \sum_{t=0}^{\infty} \beta^t u\left(c_t\right),$$

where $0 < \beta < 1$ and the utility index $u(\cdot)$ is strictly increasing and strictly concave. For simplicity, leisure is not valued.

This is a one-sector economy in which the relative price of capital in terms of consumption good is 1. Production technology is represented by a concave, homogeneous of degree one function of the capital and labor inputs:

$$Y(t) = F\left(K_t,\, n_t\right).$$

Then the central planner's problem is:

$$V(K_0) = \max_{\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t u\left(c_t\right) \right\}$$

$$\text{s.t. } \; c_t + K_{t+1} = F\left(K_t,\, n_t\right), \forall t.$$

The solutions to this problem are the Pareto optimal allocations. Then suppose we have an allocation $\left\{c_t^*,\, K_{t+1}^*,\, n_t\right\}_{t=0}^{\infty}$ solving this planner's problem and we want to support it as a competitive equilibrium. Then we need to show that there exist sequences $\{p_t^*\}_{t=0}^{\infty}$, $\{R_t^*\}_{t=0}^{\infty}$, $\{w_t^*\}_{t=0}^{\infty}$ such that:

(i) $\left\{c_t^*,\, K_{t+1}^*,\, n_t\right\}_{t=0}^{\infty}$ maximizes consumer's utility subject to the budget constraint determined by $\{p_t^*,\, R_t^*,\, w_t^*\}_{t=0}^{\infty}$.

(ii) $\{K_t^*,\, n_t\}_{t=0}^{\infty}$ maximize firm's profits.

(iii) Markets clear (the allocation $\left\{c_t^*,\, K_{t+1}^*\right\}_{t=0}^{\infty}$ is resource-feasible).

Remark 7.16 *Even though n_t can be treated as a parameter for the consumer's problem, this is not the case for the firms. These actually choose their amount of labor input each period. Therefore, we must make the sequence n_t part of the competitive equilibrium, and require that the wage level for each t support this as firms' equilibrium labor demand.*

A straightforward way of showing that the sequences $\{p_t^*\}_{t=0}^{\infty}$, $\{R_t^*\}_{t=0}^{\infty}$, $\{w_t^*\}_{t=0}^{\infty}$ exist is directly by finding their value. Notice that from concavity of $F(\cdot,\, \cdot)$,

$$\begin{aligned} R_t^* &= F_1\left(K_t^*,\, n_t\right), \\ w_t^* &= F_2\left(K_t^*,\, n_t\right) \end{aligned}$$

will ensure that firms maximize profits (or if you like, that the labor and capital services markets clear each period). In addition, homogeneity of degree 1 implies that these factor payments completely exhaust production, so that the consumer ends up receiving the whole product obtained from his factor supply.

Then the values of p_t^* remain to be derived. Recall the first order conditions in the planner's problem:

$$\beta^t u'(c_t^*) = \lambda_t^*,$$
$$\lambda_t^* = F_1\left(K_{t+1}^*, n_{t+1}\right)\lambda_{t+1}^*,$$

which lead to the centralized Euler equation

$$u'(c_t^*) = \beta u'(c_{t+1}^*) F_1\left(K_{t+1}^*, n_{t+1}\right).$$

Now, since λ_t^* is the marginal value of relaxing the planner's problem resource constraint at time t, it seems natural that prices in a competitive equilibrium must reflect this marginal value as well. That is, $p_t^* = \lambda_t^*$ seems to reflect the marginal value of the scarce resources at t. Replacing in the planner's Euler equation, we get that

$$F_1\left(K_{t+1}^*, n_{t+1}\right) = \frac{p_t^*}{p_{t+1}^*}.$$

Replacing by R_t^*, this reduces to

$$R_t^* = \frac{p_t^*}{p_{t+1}^*}. \tag{7.8}$$

It is straightforward to check that (7.8) is the market Euler equation that obtains from the consumer's first order conditions in the decentralized problem (you should check this). Therefore these prices seem to lead to identical consumption and capital choices in both versions of the model. We need to check, however, that the desired consumption and capital paths induced by these prices are feasible: that is, that these are market clearing prices. To that effect, recall the planner's resource constraint (which binds due to local non-satiation):

$$c_t^* + K_{t+1}^* = F\left(K_{t+1}^*, n_{t+1}\right), \ \forall t.$$

The equality remains unaltered if we premultiply both sides by p_t^*:

$$p_t^*\left[c_t^* + K_{t+1}^*\right] = p_t^* F\left(K_{t+1}^*, n_{t+1}\right), \ \forall t.$$

And summing up over t, we get:

$$\sum_{t=0}^{\infty} p_t^*\left[c_t^* + K_{t+1}^*\right] = \sum_{t=0}^{\infty} p_t^* F\left(K_{t+1}^*, n_{t+1}\right).$$

Finally, homogeneity of degree 1 of $F(\cdot, \cdot)$ and the way we have constructed R_t^* and w_t^* imply that

$$\sum_{t=0}^{\infty} p_t^* \left[c_t^* + K_{t+1}^*\right] = \sum_{t=0}^{\infty} p_t^* \left[R_t^* K_t^* + w_t^* n_t\right].$$

Therefore the budget constraint in the market economy is satisfied if the sequence $\left\{c_t^*,\, K_{t+1}^*\right\}_{t=0}^{\infty}$ is chosen when the prevailing prices are $\{p_t^*,\, w_t^*,\, R_t^*\}_{t=0}^{\infty}$.

Next we need to check whether the conditions for $\left\{c_t^*,\, K_{t+1}^*,\, n_t,\, p_t^*,\, w_t^*,\, R_t^*\right\}_{t=0}^{\infty}$ to be a competitive equilibrium are satisfied or not:

(i) *Utility maximization subject to budget constraint*: We have seen that the budget constraint is met. To check whether this is in fact a utility maximizing consumption-capital path, we should take first order conditions. But it is straightforward that these conditions lead to the Euler equation (7.8) which is met by the planner's optimal path $\left\{K_{t+1}^*\right\}_{t=0}^{\infty}$.

(ii) *Firms' maximization*: By construction of the factor services prices, and concavity of the production function, we have that $\{K_t^*,\, n_t\}_{t=0}^{\infty}$ are the firms' profit maximizing levels of factor inputs.

(iii) *Market clearing*: We have discussed before that the input markets clear. And we have seen that if the consumer's decentralized budget constraint is met, this implies that the planner's problem resource constraint is met for the corresponding consumption and capital sequences. Therefore the proposed allocation is resource-feasible.

Recall we mentioned convexity as a necessary assumption for the Second Welfare Theorem to hold.

Convexity of preferences entered our proof in that the first order conditions were deemed sufficient to identify a utility maximizing consumption bundle.

Convexity of the consumption possibilities set took the form of a homogeneous of degree one, jointly concave function F. Concavity was used to establish the levels of factor remunerations R_t^*, w_t^* that support K_t^* and n_t as the equilibrium factor demand by taking first order conditions on F. And homogeneity of degree one ensured that with R_t^* and w_t^* thus determined, the total product would get exhausted in factor payment - an application of the Euler Theorem.

7.5.2 The second welfare theorem in a 2-agent economy

We now assume an economy with the same production technology and inhabited by two agents. Each agent has preferences on infinite-dimensional consumption vectors represented by the function

$$U_i\left[(c_{it})_{t=0}^{\infty}\right] = \sum_{t=0}^{\infty} \beta_i^t u_i(c_{it}) \quad i = 1,\ 2,$$

where $\beta_i \in (0, 1)$, and $u_i(\cdot)$ is strictly increasing, concave, for both $i = 1, 2$.

For some arbitrary weights μ_1, μ_2, we define the following welfare function:

$$W\left[(c_{1t})_{t=0}^{\infty}, (c_{2t})_{t=0}^{\infty}\right] = \mu_1 U_1\left[(c_{1t})_{t=0}^{\infty}\right] + \mu_2 U_2\left[(c_{2t})_{t=0}^{\infty}\right].$$

Then the following welfare maximization problem can be defined:

$$V(K_0) = \max_{\{c_{1t}, c_{2t}, K_{t+1}\}_{t=0}^{\infty}} \left\{\mu_1 \sum_{t=0}^{\infty} \beta_1^t u_1(c_{1t}) + \mu_2 \sum_{t=0}^{\infty} \beta_2^t u_2(c_{2t})\right\}$$

$$\text{s.t. } c_{1t} + c_{2t} + K_{t+1} \leq F(K_t, n_t), \forall t,$$

where $n_t = n_{1t} + n_{2t}$ denotes the aggregate labor endowment, which is fully utilized for production since leisure is not valued.

If we restrict μ_1 and μ_2 to be nonnegative and to add up to 1 (then W is a convex combination of the U_i's), we have the *Negishi characterization*: by varying the vector (μ_1, μ_2), all the Pareto optimal allocations in this economy can be obtained from the solution of the problem $V(K_0)$.

That is, for every pair (μ_1, μ_2) such that $\mu_1, \mu_2 \geq 0$, $\mu_1 + \mu_2 = 1$, we obtain a Pareto optimal allocation by solving $V(K_0)$. Now, given any such allocation $\left(c_{1t}^*, c_{2t}^*, K_{t+1}^*\right)_{t=0}^{\infty}$, is it possible to decentralize the problem $V(K_0)$ so as to obtain that allocation as a competitive equilibrium outcome? Will the price sequences necessary to support this as a competitive equilibrium exist?

In order to analyze this problem, we proceed as before. We look for the values of $\{p_t^*, R_t^*, w_t^*\}_{t=0}^{\infty}$ and we guess them using the same procedure:

$$\begin{aligned} p_t^* &= \lambda_t^*, \\ R_t^* &= F_1(K_t^*, n_t), \\ w_t^* &= F_2(K_t^*, n_t). \end{aligned}$$

The planner's problem first order conditions yield

$$\begin{aligned} \mu_1 \beta_1^t u_1'(c_{1t}) &= \lambda_t, \\ \mu_2 \beta_2^t u_2'(c_{2t}) &= \lambda_t, \\ \lambda_t &= \lambda_{t+1} F_1(K_{t+1}, n_{t+1}). \end{aligned}$$

Does the solution to these centralized first order conditions also solve the consumers' decentralized problem? The answer is yes, and we can verify it by using $p_t = \lambda_t$ to replace in the previous expression for consumer 1 (identical procedure would be valid for consumer 2):

$$\begin{aligned} \mu_1 \beta_1^t u_1'(c_{1t}) &= p_t, \\ \mu_1 \beta_1^{t+1} u_1'(c_{1t+1}) &= p_{t+1}. \end{aligned}$$

So, dividing, we obtain

$$u_1'(c_{1t}) = \beta_1 u_1'(c_{1t+1}) \frac{p_t}{p_{t+1}}.$$

This is the decentralized Euler equation (notice that the multiplier μ_1 cancels out).

Next we turn to the budget constraint. We have the aggregate expenditure-income equation:

$$\sum_{t=0}^{\infty} p_t [c_{1t} + c_{2t} + K_{t+1}] = \sum_{t=0}^{\infty} p_t [R_t K_t + w_t n_t].$$

By homogeneity of degree 1 of $F(\cdot,\, \cdot)$, the factor remunerations defined above imply that if the central planner's resource constraint is satisfied for a $\{c_{1t},\, c_{2t},\, K_{t+1}\}_{t=0}^{\infty}$ sequence, then this aggregate budget constraint will also be satisfied for that chosen consumption-capital accumulation path.

However, satisfaction of the aggregate budget constraint is not all. We have an additional dilemma: how to split it into two different individual budget constraints. Clearly, we need to split the property of the initial capital between the two agents:

$$k_{10} + k_{20} = K_0.$$

Does k_{10} contain enough information to solve the dilemma? First notice that from the central planner's first order condition

$$\lambda_t = \lambda_{t+1} F_1 (K_{t+1},\, n_{t+1})$$

we can use the pricing guesses $R_t = F_1(K_t,\, n_t)$, $p_t = \lambda_t$, and replace to get

$$p_t = p_{t+1} R_{t+1}.$$

Therefore, we can simplify in the aggregate budget constraint

$$p_t K_{t+1} = p_{t+1} R_{t+1} K_{t+1}$$

for all t. Then we can rewrite

$$\sum_{t=0}^{\infty} p_t [c_{1t} + c_{2t}] = p_0 R_0 (k_{10} + k_{20}) + \sum_{t=0}^{\infty} p_t w_t n_t.$$

And the individual budgets (where the labor endowment is assigned to each individual) read:

$$\sum_{t=0}^{\infty} p_t c_{1t} = p_0 R_0 k_{10} + \sum_{t=0}^{\infty} p_t w_t n_{1t}, \tag{7.9}$$

$$\sum_{t=0}^{\infty} p_t c_{2t} = p_0 R_0 k_{20} + \sum_{t=0}^{\infty} p_t w_t n_{2t}. \tag{7.10}$$

Notice that none of them include the capital sequence directly, only indirectly via w_t. Recall the central planner's optimal consumption sequence for Agent 1 $\{c_{1t}^*\}_{t=0}^{\infty}$ (the one we wish to implement), and the price guesses: $\{w_t^* = F_2(K_t^*, n_t)\}_{t=0}^{\infty}$ and $\{p_t^* = \lambda_t^*\}_{t=0}^{\infty}$. Inserting these into (7.9), we have:

$$\sum_{t=0}^{\infty} p_t^* c_{1t}^* = p_0^* R_0^* k_{10} + \sum_{t=0}^{\infty} p_t^* w_t^* n_{1t}.$$

The left hand side $\sum_{t=0}^{\infty} p_t^* c_{1t}^*$ is the present market value of planned consumption path for Agent 1. The right hand side is composed of his financial wealth $p_0^* R_0^* k_{10}$ and his "human wealth" endowment $\sum_{t=0}^{\infty} p_t^* w_t^* n_{1t}$. The variable k_{10} is the adjustment factor that we can manipulate to induce the consumer into the consumption-capital accumulation path that we want to implement.

Therefore, k_{10} contains enough information: there is a one to one relation between the weight μ and the initial capital level (equivalently, the financial wealth) of each consumer. The Pareto optimal allocation characterized by that weight can be implemented with the price guesses defined above, and the appropriate wealth distribution determined by k_{10}. This is the Second Welfare theorem.

7.6 Uncertainty

The case with uncertainty is of special interest, because it raises the question of how Pareto domination should be defined. Let, as in the case above, the economy be composed of two-period-lived individuals, and let their utility functions be a special case of that considered in the dynastic model: utility is additively separable and of the expected-utility variety. I.e., as of when a person is born, his/her utility is some $u(c_y)$ plus $\beta E(u(c_o))$, where the expectation is taken over whatever uncertainty occurs in the next period. Also as in the dynastic model, let allocations be indexed by the history of shocks, z^t. Thus, with Z^t denoting the set of possible histories at t, a consumption allocation is a (stochastic) sequence $c = \left\{ \{(c_{yt}(z^t), c_{ot}(z^{t+1}))\}_{z^t \in Z^t} \right\}_{t=0}^{\infty} \cup c_{o,-1}(z_0)$.

We define feasibility as before, and for every possible history: for all z^t, $c_{yt}(z^t) + c_{o,t-1}(z^t)$ must be constrained either by endowments at (t, z^t) or by a similar requirement if there is (intertemporal) production. However, what does it mean that one feasible allocation, c^A, Pareto dominates another feasible allocation, c^B?

There are two quite different ways of defining Pareto domination. In the first definition, we require for c^A to dominate c^B that, for all (t, z^t), $u(c_{yt}^A(z^t)) + \beta E(u(c_{ot}^A(z^{t+1})|z^t) \geq u(c_{yt}^B(z^t)) + \beta E(u(c_{ot}^B(z^{t+1})|z^t)$ (and $c_{o,-1}^A(z_0) \geq c_{o,-1}^B(z_0)$), with strict inequality for some (t, z^t). In the second definition, we require that, for all t, $E(u(c_{yt}^A(z^t)) + \beta u(c_{ot}^A(z^{t+1}))|z_0) \geq E(u(c_{yt}^B(z^t)) + \beta u(c_{ot}^B(z^{t+1}))|z_0)$ (and $c_{o,-1}^A(z_0) \geq c_{o,-1}^B(z_0)$), with strict inequality for some t.

There is a sharp difference between these definitions: the first one treats cohorts born under different histories *as different individuals*, whereas the second definition defines the utility of cohort t in an ex ante sense. Thus, for illustration, imagine an endowment economy with constant per-capita endowments over time, normalized to 1. Thus, there is actually no aggregate uncertainty in the environment. Also, suppose that $\beta = 1$ for simplicity. Let c^A be an allocation where all consumers have $c_y = c_o = 1$, so that the total endowment at all times is split equally between young and old. Let c^B be an allocation where we introduce randomness: suppose that, from period 1 and on, either the young consume twice the amount of the old ($c_{yt}(z^t) = 4/3 = 2c_{o,t-1}(z^t)$, or vice versa, with a 50-50 coin flip determining which case applies. Does c^A dominate c^B? With the second definition of Pareto dominance, the answer is yes, given that u is strictly concave: introducing uncertainty must deliver lower ex-ante utility for all cohorts. Formally, we need to simply check that $u(1)+u(1) = 2u(1) > 0.5(u(2/3) + u(4/3)) + 0.5(u(2/3) + u(4/3)) = u(2/3) + u(4/3)$ for cohorts $t > 1$, which is true from strict concavity of u, and that $u(1) + u(1) = 2u(1) > u(1) + 0.5(u(2/3) + u(4/3))$ for cohort 1, which also follows from strict concavity of u.

Turning to the first definition, however, c^A does *not* Pareto dominate c^B, because for Pareto domination we would need to require that for *any* sequence of outcomes of the coin flips, the allocation without randomness be better, and clearly, it would not be (at least with limited curvature of u). In particular, for any t and z_t such that the current young is lucky (i.e., gets 2/3 of the total endowment), this young person would be worse off consuming (1,1): $u(4/3) + 0.5(u(4/3) + u(2/3) < u(1) + u(1)$, unless u has very high curvature.[7]

What is the argument for the first definition? It is that allocations which differ across realizations as of when a person is born cannot be compared based revealed-preference reasoning: noone ever has the ex-ante choice, where ex-ante refers to "prior to birth". Therefore, according to this definition, we need to remain agnostic as to how to make this comparison, and the formal implementation of this idea is to simply not ever allow one allocation to be better than another unless it is better for all realizations.[8] The second definition takes an explicit stand, one which possibly could be based on introspection: if I *could have* had a choice before being born, I would have preferred whatever is given by the ex-ante utility. However, note that it is hard to decide what such a utility measure would be; what would distinguish the evaluation of $E(u(c^A_{yt}(z^t)) + \beta u(c^A_{ot}(z^{t+1}))|z_0)$ from the evaluation of $E([u(c^A_{yt}(z^t)) + \beta E(u(c^A_{ot}(z^{t+1}))|z^t)]^\alpha|z_0)$, for example, for any α? Since there is no revealed-preference measurement of α, we could for example set it to a large positive number, in effect making the ex-ante perspective be "risk-loving", instead of assuming, as does the second definition, that α has to equal 1.

7.7 Hybrids

In this section a "hybrid model", and a variant of it, will be presented which shares features of the dynastic and the finite-life models above. The model is often referred to as the "perpetual-youth model", because the key assumption is that every period, a fraction $1-$

[7] The case of logarithmic curvature, for example, gives $\log(4/3) + 0.5(\log(4/3) + \log(2/3)) = \log((4/3) \cdot \sqrt{(8/9)}) = \log((4/9) \cdot \sqrt{8}) > 0 = 2u(1)$, since $16 \cdot 8 = 128 > 81$.

[8] Of course, the first definition *does* handle uncertainty as of the second period of people's lives, since it uses expected utility over those realizations.

ρ of the consumers die randomly and are replaced by newborns, and the death event is independently distributed across consumers of all age groups. I.e., from the perspective of any individual alive at t, the probability of survival until the next period is ρ. This simplifies the analysis relative to the overlapping-generations setting above, because it makes *all consumers have the same planning horizon*. We will also specialize preferences to the class that allows aggregation in wealth, which then makes all consumers—independently not only of their wealth levels but also of their ages—have the same marginal propensities to save and consume out of wealth. This is in sharp contrast to the overlapping-generations setting, where consumers of different ages have different propensities to save. There, if consumers live for two periods, the old have a zero savings propensity, since only the young save; in a version of the model where people live for more than one period, each age group in general must have a distinct marginal savings propensity, and this propensity typically declines with age.

The new model, however, shares important features with the overlapping-generations setting above. One of these features is that it allows for a nontrivial determination of long-run real interest rates. The second, related, feature is that it allows government budget deficits to have real effects, even when taxes are lump-sum; for a discussion of this topic, see the chapter on fiscal policy below.

7.7.1 The benchmark perpetual-youth model

We will first focus on a stationary environment, i.e., on one where prices and aggregate quantities are constant and where the economy is in a steady state. Thereafter, we consider transition paths.

Steady state

We assume that all consumers have the same income, e, accruing every period. Thus the consumer's problem is to maximize

$$\sum_{t=0}^{\infty}(\beta\rho)^t\frac{c_t^{1-\sigma}-1}{1-\sigma}$$

subject to $a_{t+1}+c_t=(R/\rho)a_t+e$ for all $t\geq 0$, with $a_0=0$: people are born without asset wealth. Here, R is the risk-free rate of return, and consumers obtain a higher return on lending (or pay it for loans), R/ρ. In other words, a lending consumer obtains a higher return than R if he survives, but loses the entire amount if he does not survive (in which he does not need the resources). Thus, in expectation the return is $\rho\cdot(R/\sigma)+(1-\rho)\cdot 0=R$. The higher return can be viewed as an efficient use of an annuity market.

The population is size one, with an age structure as follows: there is fraction $1-\rho$ of newborns, a fraction $(1-\rho)\rho$ of one-year-olds, and more generally a fraction $(1-\rho)\rho^s$ of s-year-olds. In order to determine the equilibrium interest rate R, we need to also model "asset supply". We will consider two economies, one with no production, in which total savings will have to equal zero, and one with neoclassical production.

In order to calculate total savings, let us solve the consumer's problem by solving his functional Euler equation. The sequential Euler equation reads

$$\frac{c_{t+1}}{c_t} = (\beta R)^{\frac{1}{\sigma}} ,$$

so the functional equivalent, using $a' = A + Ba$, which implies $c = (R/\rho)a + e - (A + Ba)$, reads

$$e - A + (\frac{R}{\rho} - B)(A + Ba) = (\beta R)^{\frac{1}{\sigma}} \left(e - A + (\frac{R}{\rho} - B)a \right).$$

From this it follows, by equating coefficients, that

$$B = (\beta R)^{\frac{1}{\sigma}}$$

and that

$$A = e \cdot \frac{(\beta R)^{\frac{1}{\sigma}} - 1}{\frac{R}{\rho} - 1}.$$

To find total savings, note that a consumer just having turned s years of age has accumulated $a_s = A(1 - B^s)/(1 - B)$ (if $B \neq 1$, and zero otherwise). Thus, total savings equal

$$\sum_{s=0}^{\infty} (1 - \rho)\rho^s A \frac{1 - B^s}{1 - B} = \frac{(1 - \rho)A}{1 - B} \sum_{s=0}^{\infty} \rho^s (1 - B^s) = \frac{e}{\frac{R}{\rho} - 1} \left(\frac{1 - \rho}{1 - \rho\,(\beta R)^{\frac{1}{\sigma}}} - 1 \right)$$

so long as $\beta R \neq 1$, and zero otherwise.

Turning to the equilibrium determination of R, first consider an endowment economy. Since total savings now have to equal zero, we see that $\beta = 1$ must hold; no other value for R than $1/\beta$ makes total savings zero. Thus, in an endowment economy where consumers have constant endowments, the interest rate has to equal the subjective discount rate, just like in the dynastic model. Other endowment structures can deliver other outcomes, however.

In the neoclassical growth context, say with a production function of the $k^\alpha n^{1-\alpha}$ variety where each consumer has one unit of labor per period, a steady state would make $e = (1 - \alpha)k^\alpha$ (since $n = 1$ in equilibrium) and $R = \alpha k^{\alpha - 1} + 1 - \delta$. Thus, requiring that total savings equal k, we must have

$$\left(\frac{R - 1 + \delta}{\alpha} \right)^{\frac{1}{\alpha - 1}} = \frac{e}{\frac{R}{\rho} - 1} \left(\frac{1 - \rho}{1 - \rho\,(\beta R)^{\frac{1}{\sigma}}} - 1 \right).$$

This equation determines the equilibrium interest rate R. Since capital has to be positive, we see that $\beta R > 1$ must follow; otherwise the term in parenthesis on the right-hand side is negative. The interpretation of this condition is that, since consumers need to hold capital, the interest rate has to rise relative to $1/\beta$ to induce people to accumulate so that the capital stock can be held. We also note that $\beta R < \rho^{-\sigma}$ must hold. It is straightforward to show that there is a solution R to this equation such that $\beta R \in (1, \rho^{-\sigma})$.[9]

[9] The demonstration that R only has one solution in this interval should be possible too; this task remains to be completed. . . .

Transition

Along a transition path, it is possible to derive a second-order difference equation for k_t. Here, note that although $c_{t+1} = (\beta R)^{\frac{1}{\sigma}} c_t$ holds for all individuals born at t and surviving until $t+1$, we will not have that average consumption at $t+1$ will be $(\beta R)^{\frac{1}{\sigma}}$ times consumption today: it will be lower than that, because those consumers who die are replaced by poorer ones (given that consumption paths are rising over time). To derive the difference equation, first ... (details left out, since this will be a "homework").

7.7.2 Introducing a life cycle

It is possible to use the same structure to introduce life-cycle patterns, for example via earnings. Say that, upon birth, people face a constant probability, $1 - \hat{\rho}$, of becoming "old", and upon becoming old they face a constant probability, $1 - \rho$, of death. Thus, the second phase of their life is just like for consumers in the model just described, but the first phase is now different. Supposing, thus, that consumers have endowments e_y in the first phase of their life and e_o in their second, we can solve their consumption-saving problems at different ages and with different wealth holdings using recursive methods as follows ... (details left out, since this will be another "homework").

Chapter 8

Consumption and labor choice

We begin the more applied part of the text to discuss how consumption and labor supply are chosen. These are, arguably, two of the most important choices households make, so they deserve special focus for this reason. They are also integral parts of most analyses in macroeconomics. We will focus on the dynastic household, but much of the insights here carry over to households with finite lifetimes. Some of the discussion will involve some details of the market structure, but the general purpose is to focus on individual choice both in the very long run and in response to short-term fluctuations. Thus, this chapter prepares for, and offers intuition behind, the analyses in the growth and business-cycles chapters that come next.

We begin with a static model. The idea here is that many aspects of labor supply can be understood in such a setting. Moreover, the static model can be closely tied to a dynamic model under balanced growth. We then introduce dynamics and illustrate with simple examples without uncertainty.

8.1 A static model

In a frictionless market with one consumption good and where leisure is valued, consumption and hours worked, along with the wage, can be viewed to be determined by labor demand and labor supply. Thus consider a static economy where the resource constraint is given by $c = F(k,n) - \delta k$, where k is capital and n is labor. In the static economy, we regard k as exogenous, and notice that we specify production by a function F before and explicitly include a depreciation cost for capital the way it would appear in a steady state where capital is held constant. A representative consumer enjoys consumption and leisure according to a function $u(c,l)$, where l is leisure; l and n sum up to a total time endowment, assumed to be 1 for simplicity. The social planner would thus simply maximize $u(F(k,1-l),l)$ by choice of l. This choice results in the first-order condition

$$\frac{u_l(F(k,1-l)-\delta k,l)}{u_c(F(k,1-l)-\delta k,l)} = F_n(k,1-l).$$

This equation solves for l given k and it has the interpretation that the marginal rate of substitution between labor and leisure has to equal the marginal rate of transformation between labor and consumption.

In a decentralized economy we can talk about labor demand and labor supply. The obvious decentralized market structure here is one with competitive markets for inputs and output. Firms maximize $F(k,n) - rk - wn$ by choice of k and n; this allows us to talk about each firm's demand for capital and labor. We will as usual assume that F has constant returns to scale (and is concave, so that first-order conditions can be used). An individual firm will set $F_k(k,n) = r$ and $F_n(k,n) = w$, in principle determining k and n as a function of the prices r and w. For general values for r and w, the firm decision can lead to either zero production, be ill-defined (because if r and w are such that the firm can make profits, there is no bound to how high the profits can become since the firm can scale production arbitrarily), or lead to an indeterminate solution (in the knife-edge case). However, since capital, unlike labor, is exogenous here, we can think of equilibrium r as determined residually from the firm's first-order condition for capital—it becomes whatever it has to become as a function of the outcome for the labor input. Therefore, we can think of aggregate labor demand as being determined by the firm's first-order condition for labor:

$$F_n(k,n) = w.$$

This expression gives the inverse labor demand function; solving instead for n as a function of w (and k) we obtain the labor demand function.

What does the labor demand function look like? The answer is entirely dependent on the shape of the production function. Macroeconomists almost always use the Cobb-Douglas formulation: $F(k,n) = Ak^{\alpha}n^{1-\alpha}$ where A and α are given parameters. The reason, as discussed in previous chapters, is that the labor share—total wages as a fraction of output—has remained remarkable constant in the U.S. for a long time so that one might want to require an F whose value for $F_n(k,n)n/F(k,n) = F_n(k/n,1)/F(k/n,1)$ is indeed equal to a constant (here $1-\alpha$) for all values of k/n, since k/n has grown a lot over the same period. For the Cobb-Douglas function, thus, we obtain labor demand as $n = [A(1-\alpha)]^{\frac{1}{\alpha}} k w^{-\alpha}$, an isoelastic function shifted one-for-one by the level of capital. Two remarks are worth making here, however. First, the labor share has moved significantly over time in some economies and seems to have displayed a downward trend recently, since about 2000, in many countries (including the U.S.). Second, it is not impossible to formulate a production function outside the Cobb-Douglas class that can deliver constant shares. Consider a CES: $y = (\alpha(A_k k)^{\rho} + (1-\alpha)(A_n n)^{\rho})^{1/\rho}$, where ρ regulates the substitutability between the inputs ($\rho = -\infty$ is Leontief, $\rho = 0$ is Cobb-Douglas, and $\rho = 1$ is perfect substitutes) and there are now two separate technology variables, one capital-augmenting and one labor-augmenting. As was pointed out in the growth chapter, we can have a balanced growth path for this formulation, regardless of the value of ρ, with y and k growing at the same rate and n staying constant if A_k is constant but A grows at the same rate as output. Thus, consider a CES function with $\rho \neq 1$ and such that A_n is growing at a roughly constant rate; then it would not be surprising to observe some movements in the labor share around a stationary value as k and n fluctuate, for whatever reason, around their their balanced paths. Supposing that $\rho < 0$ ($\rho > 0$), a declining labor share could then be interpreted as a falling (rising) value of $(A_k/A_n)(k/n)$.[1] Regardless of the degree of substitutability, for a given labor input the wage will depend greatly on the development of technology. The nature of technological change—which factor(s) is augmented—will be important for the wage too if there is more or less substitutability than Cobb-Douglas.

Turning to labor supply, the consumer would maximize $u(c, l)$ subject to $c = w(1-l) + (r-\delta)k$ by choice of c and l. The outcome is a first-order condition that reads $u_c(w(1-l)+(r-\delta)k, l)w = u_l(w(1-l)+(r-\delta)k, l)$. This equation implicitly defines labor supply: $n = 1-l$ as a function of w. Not surprisingly, the shape of the utility function is thus a key determinant of labor supply. Like in the case for labor demand, one can restrict the functional form based on some long-run facts. The key long-run fact on the labor-supply side is that working hours per household have remained roughly constant over a long period of time, despite massive increases in the wage. Intuitively, it must therefore be the case that we need a utility function such that the income and substitution effects go in opposite directions and, in fact, cancel. The increasing wage will induce people to work more, since it pays off more and more to work, but for a given amount of hours worked the higher wage causes a higher demand for leisure (if leisure is a normal good) going in the opposite direction. A utility function that has the desired property is (any monotone transformation of) $cv(l)$; assume also that v is not only strictly increasing but that $cv(l)$ is strictly quasiconcave. That these preferences will be necessary to match the long-run facts will be proved in Section 8.2.1 below.

To illustrate how the household's problem is solved using this functional form, note that the first-order condition becomes $v(l)w = cv'(l)$. Let us first simplify the discussion and suppose there is no capital income. Then, with the budget substituted in, one sees that w cancels and that one obtains $v(l) = (1-l)v'(l)$, an equation that determines l based on the shape of v only: income and substitution effects cancel. If there is capital income, the first-order condition becomes $v(l) = (1-l+(r-\delta)k/w)v'(l)$. Now an increased wage level (it is straightforward to verify it) will increase the amount of working time supplied to the market if the added wealth is positive—if $(r-\delta)k > 0$). The intuitive reason is that the income effect of an increased amount of wealth is smaller when there is other income, so the substitution effect will dominate.[2] In the long run, of course, we know from the Kaldor facts discussed in the growth section that k will grow at the same rate that w will grow, and r will be constant, so the fact that labor supply is constant is consistent with the presence of capital income: rk/w will matter for the determination of l but this variable will have no trend and therefore the outcome for l will be stationary.

Putting together labor demand and labor supply and eliminating prices it is straightforward to see that we reproduce the planner's first-order condition for l. However, the purpose of looking separately at demand and supply was to gain some intuitive insights on the firm

[1] The labor share equals $(1-\alpha)/(\alpha(A_k k/(A_n n))^\rho + 1 - \alpha)$.

[2] If the added wealth is negative, then the income effect is stronger than before and an increased wage will lead to less work.

and consumer levels into the determinants not only of hours worked but also of wages. Given the specialized preferences we use, in particular, we see that the constancy of labor supply will imply, under Cobb-Douglas production, that the wage will be proportional to Ak^{α} and hence follow its trend.

8.2 A dynamic model

We restrict attention to an economy with dynastic households (a restriction that will not be important for the main points here). Thus we describe preferences by $\sum_{t=0}^{\infty} \beta^t u(c_t, l_t)$. We also use a neoclassical setting so the resource constraint simply reads $c_t + k_{t+1} = F(k_t, A_t(1-l_t)) + (1-\delta)k_t$. We will establish that a certain utility function is necessary and sufficient for exact balanced growth where the Kaldor facts are satisfied.[3] We will then look at several specific cases in order to gain intuition for how consumption and saving will be determined in dynamic economies.

8.2.1 Preference requirements for exact balanced growth

Let us now look more carefully at the preference requirements necessary for balanced growth, i.e., a formulation such that if labor-augmenting productivity grows at a constant rate it is not only feasible but also desirable to have consumption growth at that same rate but also a constant labor supply. These preferences will exhibit a constant rate of intertemporal substitution regarding consumption (or a consumption-leisure bundle) and, regarding leisure, will require that income and substitution effects cancel. This class of utility functions is given by a $\sigma \geq 0$ and a v such that

$$u(c, l) = \lim_{s \to \sigma} \frac{(cv(l))^{1-s} - 1}{1 - s}.$$

There are many commonly used special cases. One is $\theta \log c + (1-\theta) \log l$. Another (more general) formulation, using labor input instead as leisure, is $\log c - B\frac{n^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$. As we shall see, the former places rather strong restrictions on curvature, whereas the second one has curvature that is parameterized by γ. We will mainly use the latter in our illustrations.

Constant elasticity of intertemporal substitution

We will proceed by abstracting first from valued leisure and thus have $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$. Before proceeding to the argument why this function is needed, let us interpret it. The *elasticity of intertemporal substitution* between c_t and c_{t+k} is defined in terms of consumption choice as

$$\left. \frac{d \log \frac{c_{t+k}}{c_t}}{d \log R_{t,t+k}} \right|_{\bar{U}},$$

[3]That technology has to be labor-augmenting is something we will discuss and verify in detail in Chapter 9 below. This discussion involves only restrictions on technology and thus does not overlap with the discussion here.

where the notation $R_{t,t+k}$ is used to represent the relative price between the two consumption goods, i.e., the accumulated gross interest rate across the k periods and $\bar{U}$ indicates a substitution along an indifference curve, i.e., taking the utility level $\bar{U}$ as given (a "compensated" elasticity). That is, if you change the gross interest rate by 1 percent, how will the ratio of consumption levels at the two dates change (also in percent)? For the particular functional form here we will see that this elasticity is constant, independent of consumption levels. The marginal rate of substitution—which will equal the relative price—is $\frac{u'(c_t)}{\beta u'(c_{t+k})} = \left(\frac{c_{t+k}}{c_t}\right)^{\sigma} \beta^{-1}$ so for this function it is possible to write the ratio as a function of the marginal rate of substitution (or the price): $\frac{c_{t+k}}{c_t} = MRS_{t,t+k}^{\frac{1}{\sigma}} \beta^{\frac{1}{\sigma}}$ as we move along an indifference curve. Given the simple relation we have that the sought-after elasticity is constant and equal to $1/\sigma$.

It is perhaps useful to pay extra attention to the logarithmic special case because it is commonly used and has a special feature: under logarithmic preferences, income and substitution effects of an interest-rate change cancel. With logarithmic preference, the marginal rate of substitution between the two goods involved will be proportional to the ratio consumed of the two goods, and thus the relative income shares will remain constant (e.g., the ratio of c_t and c_{t+1}/R_{t+1} will be constant and equal to $1/\beta$).

The case without valued leisure

We now proceed to argue why we need a constant elasticity of intertemporal substitution in order for consumers to choose balanced consumption growth. The Euler equation has to hold for all t, or equivalently for all starting levels c_t. Thus we need to require $u'(c)/u'(gc)$ to be constant for all c. Denoting this constant $\#$, and recognizing that g would have to depend on $\#$, we have $u'(c) = u'(g(\#)c)\#$ for all c. This allows us to differentiate with respect to c to obtain $u''(c) = u''(g(\#)c)\#g(\#)$. Dividing the first equation by the second, we obtain $\frac{u''(c)}{u'(c)} = \frac{u''(g(\#)c)g(\#)}{u'(g(\#)c)}$. Multiply by c on both sides to obtain $\frac{u''(c)c}{u'(c)} = \frac{u''(g(\#)c)g(\#)c}{u'(g(\#)c)}$. Since this holds for all c and $g(\#)$ is an arbitrary positive number, this means that $\frac{u''(c)c}{u'(c)}$ is independent of c, i.e., the expression must equal a constant. Denote the constant a. Thus we have $u''(c)/u'(c) = a/c$. The "trick" here is to see that this can be written as $d \log u'(c)/dc = a \cdot d(\log c + b)/dc$ for any arbitrary constant b. Developing the expression slightly, we have $d \log u'(c)/dc = d(\log c^a + ab)/dc$. This gives that $\log u'(c) = \log c^a + B$, where B is a constant. Thus, $u'(c) = Ac^a$ for some constant A. This means that $u(c)$ is of the functional form stated. (Notice that the case $\sigma = 1$ is subsumed here—it corresponds to $a = -1$.) Restrictions of course need to be placed jointly on A and a so that u is strictly increasing and strictly concave, which leads to the formulation adopted.

Valued leisure

In the case without leisure, we derived above that $u'(c)$ has to be of the form Ac^a. What this means in a context with leisure is that $u'(c)$ must be of the form $A(l)c^{a(l)}$: l is constant on a balanced path, and thus can be an argument of any constant appearing in $u'(c)$. This means that $u(c, l)$ must be of the form $B(l)c^{b(l)} + D(l)$ or, if $a(l) = -1$, $B(l) \log c + D(l)$.

However, we need to make sure that the first-order condition for labor is satisfied along a balanced path. Taking our functional form, and replacing w by a constant, e, times c—since they need to grow at the same rate—we obtain, for $b(l) \neq 0$,

$$u_c(c,l)w = B(l)b'(l)c^{b(l)-1}ec = u_l(c,l) = B'(l)c^{b(l)} + B(l)(\log c)b'(l)c^{b(l)} + D'(l),$$

an expression that needs to be met for all c. Because it needs to hold for all c, it is clear that unless $b(l) = 0$ (the log case), $b'(l) = D'(l) = 0$ has to hold, allowing us to conclude that for a balanced growth path with l constant and equal to an arbitrary value within some given bound, $b(l)$ and $D(l)$ have to be constants. In the log case, we obtain a similar equation where $b'(l) = 0$ is still needed for the equation to hold for all c but where $D(l)$ can be a function that depends on l; however, now $B'(l)$ must be zero so $B(l)$ must be a constant. Thus, we are left with a utility function $B(l)c^b$, for $b \neq 0$, or $B \log c + D(l)$. This completes the argument.

It is perhaps instructive to point out that some commonly used utility functions do not admit balanced growth with constant labor supply. One is the case $\frac{c^{1-\sigma}-1}{1-\sigma} + B\frac{l^\psi}{\psi}$ for $\sigma \neq 1$; additivity only works if σ is equal to one (the logarithmic case). A second case is the so-called GHH utility function (Greenwood, Hercowitz, and Huffman, xyz): $\frac{(c+v(l))^{1-\sigma}-1}{1-\sigma}$. This formulation amounts to there being no wealth/income effect of a wage change. I.e., in the first-order condition for labor vs. leisure, consumption drops out and labor supply becomes a(increasing) function of the wage only. Clearly, an ever-increasing wage would then lead to ever-increasing labor supply. It should be noted that for any utility function that does not match the long-run facts it would be possible to restore the facts by introducing exogenous trend factors in the utility function. For example, in the GHH case, if one assumes $\frac{(c+B_t v(l))^{1-\sigma}-1}{1-\sigma}$, where B_t is shifting up at exactly the rate of consumption growth, the consumer would choose constant labor supply. However, such a formulation would call for a deeper explanation of the increased value of leisure, and the results would not be robust if consumption growth were to change for technological reasons.

8.2.2 Labor-leisure and consumption choice over time without saving

For the rest of the chapter, we abstract from long-run growth (for notational convenience only) but allow time-varying labor-augmenting technology. The idea is now to look at a number or interesting cases in order to build intuition for how consumption and labor are determined in dynamic models. We will look at models where markets work, so that the planning problem can be analyzed.

It is straightforward to formulate the planning problem and derive first-order conditions: an (intertemporal) Euler equation and a labor-leisure tradeoff (which is intratemporal). Such a model would, in general, have nontrivial transition dynamics for saving and hours worked. The purpose is not to emphasize these, let alone characterize them in full generality, but rather to emphasize some mechanisms. For this, let us consider some extreme special cases that serve an illustrative purpose. The first case is trivial: if capital plays no role in production, i.e., $y_t = A_t n_t$ for all t, so that there is really no possibility to move resources across time, the model reduces to a static one without capital. Thus, no matter how much labor-augmenting technology moves over time, the labor supply will be constant. The reason is that income and substitution effects cancel: movements in A amount to movements in the wage, and because there is no other income than labor income, labor supply will be chosen to be constant and satisfying $v(l) = (1-l)v'(l)$ (where $n = 1 - l$).

8.2.3 A linear savings technology

In a second example, suppose instead that $F(k_t, A_t n_t) = bk_t + A_t n_t$, so that capital is productive as well and there are no decreasing returns to either capital or labor. Now capital accumulation is possible at the gross rate $b + 1 - \delta$ between any two periods. Here, the planning problem will look identical to a consumer problem where the prices on capital and labor, r_t and w_t, from the firm's first-order conditions, are simply b and A_t, respectively. To simplify the problem further, assume that $\beta(b + 1 - \delta) = 1$. The consumer's problem then reads

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left(\log c_t - B \frac{n_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \right) \quad \text{s.t.}$$

$$c_t + k_{t+1} = w_t n_t + k_t/\beta \quad \text{and } .$$

Here, to simplify even more, we will suppose that $k_0 = 0$ so that all resources available for consumption derive from working during the consumer's lifetime. We will also allow capital to be negative so as to illustrate how unrestricted movements of resources over time interact in an important way with the labor-supply decision.[4] The maintained formulation can also be viewed as an open-economy interpretation where the gross international interest rate is constant and equal to $1/\beta$.

The Euler equation for this problem implies that $c_t = c_{t+1} \equiv c$ for all t, i.e., that there will be complete consumption smoothing. The labor-leisure first-order condition then reads

$$\frac{w_t}{c} = B n_t^{\frac{1}{\gamma}}.$$

The resource constraint can be expressed as a discounted sum and the resulting expression can be written $c/(1-\beta) = \sum_{t=0}^{\infty} \beta^t w_t n_t$.

We shall proceed toward solving this problem momentarily but before doing this we can already note that a very different result than in the static model is expected here. Suppose time periods are quite short (i.e., β is very close to 1), so that a wage change in a given period does not influence present-value income, and hence c, much at all. Then it follows from the first-order condition above that a wage change will change hours worked rather directly. Thus, the income effect of a higher wage in a given period is thus very limited. Moreover, the effect of wages on labor supply can be strong, namely if γ is high. Intuitively,

[4] A no-Ponzi-game restriction is assumed as well.

in this model there is intertemporal substitution of hours worked: if the wage is high at one point in time and low at another point in time the consumer works more in the former and less in the latter, while moving income across period using borrowing and lending.

From inserting the first-order condition into the resource constraint and simplifying we obtain a solution for consumption:

$$c = B^{-\gamma}(1-\beta)\sum_{t=0}^{\infty}\beta^t w_t^{1+\gamma}$$

Thus, suppose compare different wage paths such that $\sum_{t=0}^{\infty}\beta^t w_t^{1+\gamma}$ is constant. Then consumption will not change in any period. For example, suppose the wage in period t_1 goes up marginally and the wage in some other period t_2 falls so as to keep $\beta^{t_1}w_{t_1}+\beta^{t_2}w_{t_2}$ constant. Then, since c will not change, the effect of the wage changes on labor supply, expressed in an elasticity form and for each of the two time periods, is $d\log n/d\log w=\gamma$.

Notice also that if we were to impose a "borrowing constraint", so that capital holdings could not fall below a certain value, then for a period in which this constraint binds—when the consumer would like to increase consumption by borrowing more—a change in the wage will have a very different effect. An increased wage would have a much stronger, positive effect on consumption and hence working hours would barely change, at least for a small enough wage increase. Similarly, a fall in the wage would make consumption fall further and have a very small effect on labor supply.

Motivated by the example above, let us finally consider a formal definition of labor-supply elasticity: the percentage change in labor supply from a one-percent increase in the wage keeping the marginal utility of wealth constant. This concept was proposed by Ragnar Frisch and is usually referred to as the Frisch elasticity. The marginal utility of wealth can be thought of in terms of goods available for consumption in any period here, and the marginal utility of such resources will equal the marginal utility of consumption. In a model where consumption and leisure are separable in utility, holding the marginal utility of wealth constant is thus equivalent to holding consumption constant. It is straightforward to see that with the preferences assumed here, the Frisch elasticity becomes precisely γ. Consider, however, the commonly used $u(c,l)=\theta\log c+(1-\theta)\log l$: what is the Frisch elasticity in this model? The first-order condition in any period reads

$$\frac{\theta w_t}{c_t} = -\frac{1-\theta}{1-n_t}.$$

To obtain the Frisch elasticity, take log on both sides and use the fact that consumption is constant so that $d\log(1-n_t)=-d\log w_t$. Because $d\log(1-n)=-(d\log n)n/(1-n)$ we thus obtain a Frisch elasticity equal to $(1-n)/n$. That is, the elasticity depends on the level of labor supply (relative to the amount of leisure). In typical calibrations, leisure is usually thought of 2/3 of the total available time and hours worked as 1/3. This implies a Frisch elasticity of 2. This number is very high compared to most microeconomic estimates, which tend to range between 0 and 1/2. For a recent survey and view of the literature, see Chetty (2009).

8.2.4 Decreasing marginal returns to savings

Suppose instead we use a more empirically reasonable production function such as the Cobb-Douglas function. Then labor income cannot be generated and transformed linearly across periods on the level of the whole economy so intertemporal substitution of leisure is not as straightforward. First, there are decreasing returns to working in a given period, and second, there are decreasing returns to saving the working income. So if one asks about the effects of, say, an increase in the level of technology in period t on hours worked, just how does the answer differ from that obtained above? Maintaining the utility function $\log c - B\frac{n^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$, we obtain first-order conditions as follows:

$$\frac{(1-\alpha)k_t^\alpha A_t^{1-\alpha} n_t^{-\alpha}}{c_t} = B n_t^{\frac{1}{\gamma}}.$$

$$\frac{1}{c_t} = \beta \frac{\alpha k_{t+1}^{\alpha-1}(A_t n_t)^{1-\alpha} + 1 - \delta}{c_{t+1}}.$$

To make the most extreme assumption of decreasing returns to capital within this framework, suppose $\delta = 1$, so that capital depreciates fully after use; this assumption is not reasonable for short time horizons but serves as illustration. Then it is straightforward to show that these equations, together with the resource constraint, imply that $k_{t+1} = \alpha\beta k_t^\alpha (A_t n_t)^{1-\alpha}$, i.e., that the rate of saving is constant and equal to $\alpha\beta$ (this guess was verified to solve the Euler equation in earlier chapters and still works here) and that the first-order condition for leisure simplifies to

$$\frac{1-\alpha}{n_t} = B n_t^{\frac{1}{\gamma}}.$$

Hence, n_t becomes constant and independent of both A_t and k_t! Decreasing returns are, apparently, strong enough in this case to totally offset any desire to intertemporally substitute labor efforts. Increased productivity thus leads to increased production and consumption at all dates (though less so further into the future) but no changes in hours worked at any date. Here (and in general), the amount of hours worked depend not just on the utility function but on the production technology, though only on its elasticity with respect to labor: the higher this elasticity, the higher is working effort.

With less than full depreciation, capital accumulation becomes "more linear" and there will be intertemporal substitution of labor.

Chapter 9

Growth

Growth is a vast subject within, and its first-order aim is to explain basic facts about the long-term behavior of different economies. The current chapter is an introduction to this subject, and it is divided into three sections. In the first section, we set forth the motivation for the theory: the empirical regularity which it seeks to explain. The second section is about exogenous growth models, i.e., models in which an exogenous change in the production technology results in income growth as a theoretical result. Finally, the third section introduces technological change as a decision variable, and hence the growth rate becomes endogenously determined.

9.1 Some motivating long-run facts in macroeconomic data

9.1.1 Kaldor's stylized facts

The first five "facts" refer to the long-run behavior of economic variables in an economy, whereas the sixth one involves an inter-country comparison.

1) The growth rate of output g_y is roughly constant over time.

2) The capital-labor ratio $\frac{K}{L}$ grows at a roughly constant rate.

3) The capital-income ratio $\frac{K}{y}$ is roughly constant (presumes that capital is measured as accumulated foregone consumption).

4) Capital and labor shares of income are close to constant.

5) Real rates of return are close to constant.

6) Growth rates vary persistently across countries.

9.1.2 Other facts

In addition to these classical facts, there are also other empirical regularities that guide our study of long-run growth. These are:

1) $\frac{Y}{L}$ is very dispersed across countries; a factor of over 30 separates the richest country from the poorest country.

2) The distribution of $\frac{Y}{L}$ does not seem to spread out (although the variance has increased somewhat, but then moving mass toward the corners).

3) Countries with low incomes in 1960 did not show on average higher subsequent growth (this phenomenon is sometimes referred to as "no absolute (β) convergence").

4) There is "conditional convergence": within groups classified by 1960 human capital measures (such as schooling), 1960 savings rates, and other indicators, a higher initial income y_0 (in 1960) was positively correlated with a lower growth rate g_y. This is studied by performing the "growth regression":

$$g_{y,\,i}^{1960-1990} = \alpha + \beta \log y_{0i} + \gamma \log edu_{0i} + \varepsilon_i, \quad i = 1, \ldots, n.$$

Then controlling for the initial level of education, the growth rate was negatively correlated with initial income for the period 1960-1990: $\widehat{\beta} < 0$. If the regression is performed without controlling for the level of education, the result for the period is $\widehat{\beta} = 0$, i.e., no absolute convergence, as mentioned above.

5) Growth in factor inputs (capital, labor) does not suffice in explaining output growth. The idea of an "explanation" of growth is due to Solow, who envisaged the method of "growth accounting". Based on a neoclassical production function

$$y = zF(K, L),$$

the variable z captures the idea of technological change. If goods production is performed using a constant-returns-to-scale technology, operated under perfect competition, then (by an application of the Euler Theorem) it is possible to estimate how much out of total production growth is due to each production factor, and how much to the technological factor z. The empirical studies have shown that the contribution of z (the Solow residual) to output growth is very significant.

6) In terms of raw correlations, and partial correlations, there are many findings in the growth-regression literature; to mention a few often-discussed variables, output growth correlates positively with the growth of the foreign trade volume and measures of human capital (education levels), and output per capita correlates positively with investment rates and measures of openness and negatively with population growth rates.

7) Workers of all skill classes tend to migrate to high-income countries.

We will revisit these facts below in various ways.

9.2 Growth theory I: optimal steady states and convergence

We will now study, in more detail, the model where there is only one type of good, that is, only one production sector: the one-sector optimal growth model. This means that we will revisit the Solow model under the assumption that savings are chosen optimally. Will, as in Solow's model, output and all other variables converge to a steady state? It turns out that the one-sector optimal growth model does produce global convergence under fairly general conditions, which can be proven analytically. If the number of sectors increases, however, global convergence may not occur. However, in practical applications, where the parameters describing different sectors are chosen so as to match data, it has proven difficult to find examples where global convergence does not apply.

We thus consider preferences of the type

$$\sum_{t=0}^{\infty} \beta^t u\left(c_t\right)$$

and production given by

$$c_t + k_{t+1} = f(k_t),$$

where

$$f(k_t) = F\left(k_t,\ N\right) + \left(1 - \delta\right) k_t$$

for some choice of N and δ (which are exogenous in the setup we are looking at). Under standard assumptions (namely strict concavity, $\beta < 1$, and conditions ensuring interior solutions), we obtain the Euler equation:

$$u'\left(c_t\right) = \beta u'\left(c_{t+1}\right) f'\left(k_{t+1}\right).$$

A *steady state* is a "constant solution":

$$\begin{aligned} k_t &= k^* \ \forall t \\ c_t &= c^* \ \forall t. \end{aligned}$$

This constant sequence $\{c_t\}_{t=0}^{\infty} = \{c^*\}_{t=0}^{\infty}$ will have to satisfy:

$$u'\left(c^*\right) = \beta u'\left(c^*\right) f'\left(k^*\right).$$

Here $u'\left(c^*\right) > 0$ is assumed, so this reduces to

$$\beta f'\left(k^*\right) = 1.$$

This is the key condition for a steady state in the one-sector growth model. It requires that the gross marginal productivity of capital equal the gross discount rate $(1/\beta)$.

Suppose $k_0 = k^*$. We first have to ask whether $k_t = k^* \ \forall t$ - a solution to the steady-state equation - will solve the maximization problem. The answer is clearly yes, provided that both the first order *and* the transversality conditions are met. The first order conditions are met by construction, with consumption defined by

$$c^* = f(k^*) - k^*.$$

The transversality condition requires

$$\lim_{t \to \infty} \beta^t F_1 [k_t, \ k_{t+1}] \, k_t = 0.$$

Evaluated at the proposed sequence, this condition becomes

$$\lim_{t \to \infty} \beta^t F_1 [k^*, \ k^*] \, k^* = 0,$$

and since $F_1 [k^*, \ k^*] \, k^*$ is a finite number, with $\beta < 1$, the limit clearly is zero and the condition is met. Therefore we can conclude that the stationary solution $k_t = k^* \ \forall t$ does maximize the objective function. If f is strictly concave, then $k_t = k^*$ is the unique strictly positive solution for $k_0 = k^*$. It remains to verify that there is indeed one solution. We will get back to this in a moment.

Graphically, concavity of $f(k)$ implies that $\beta f'(k)$ will be a positive, decreasing function of k, and it will intersect the horizontal line going through 1 only once as can be seen in Figure 9.1.

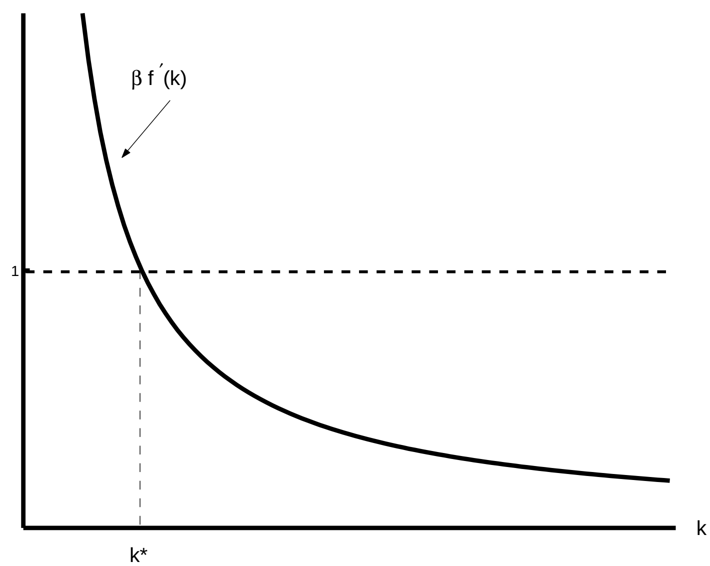

Figure 9.1: The determination of steady state

9.2.1 Properties of the capital accumulation function

Capital accumulation is given by $k' = g(k)$. In order to characterize the path of capital accumulation, it is therefore important to characterize g as much as possible. The present section presents a sequence of results on g, all of which are implications from the dynamic-programming analysis. These results are interesting from various perspectives. Taken together, they also have implications for (global) convergence, which will be discussed in the following section.

Throughout, we will use the following assumptions on primitives:

(i) u and f are strictly increasing, strictly concave, and continuously differentiable.

(ii) $f(0) = 0$, $\lim_{k \to 0} f'(k) = \infty$, and $\lim_{k \to \infty} f'(k) \equiv b < 1$.[1]

(iii) $\lim_{c \to 0} u'(c) = \infty$.

(iv) $\beta \in (0,\ 1)$.

We thus have the following problem:

$$V(k) = \max_{k' \in [0, f(k)]} \left\{ u\left[f(k) - k'\right] + \beta V(k') \right\},$$

leading to $k' = g(k)$ satisfying the first-order condition

$$u'\left[f(k) - k'\right] = \beta V'(k').$$

Notice that we are assuming an interior solution. This assumption is valid since assumptions (ii), (iii), and (iv) guarantee interiority.

Properties of $g(k)$:

(i) $g(k)$ is single-valued for all k.

This follows from strict concavity of u and V (recall the theorem we stated previously) by the Theorem of the Maximum under convexity.

(ii) $g(0) = 0$.

This follows from the fact that $f(k) - k' \geq 0$ and $f(0) = 0$.

(iii) There exists $\bar{k}$ s.t. $g(k) \leq \bar{k}$ for all $k < \bar{k}$. Moreover, $\bar{k}$ exceeds $(f')^{-1}(1/\beta)$.

The first part follows from feasibility: because consumption cannot be negative, k' cannot exceed $f(k)$. Our assumptions on f then guarantee that $f(k) < k$ for high enough values of k: the slope of f approaches a number less than 1 as k goes to infinity. So $g(k) < k$ follows. The characterization of $\bar{k}$ follows from noting (i) that $\bar{k}$ must be above the value that maximizes $f(k) - k$, since $f(k)$ is above k for very small values of k and f is strictly concave and (ii) that therefore $\bar{k} > (f')^{-1}(1) > (f')^{-1}(1/\beta)$.

[1] It is not necessary for the following arguments to assume that $\lim_{k \to 0} f'(k) = \infty$. They would work even if the limit were strictly greater than 1.

(iv) $g(k)$ is continuous.

This property, just as Property 1, follows from the Theorem of the Maximum under convexity.

(v) $g(k)$ is strictly increasing.

We argued this informally in the previous section. The formal argument is as follows.

Proof. Consider the first-order condition:

$$u'\left[f(k)-k'\right]=\beta V'(k').$$

$V'(\cdot)$ is decreasing, since $V(\cdot)$ is strictly concave due to the assumptions on u and f. Define

$$\begin{aligned} LHS\,(k,k') &= u'\left[f(k)-k'\right] \\ RHS\,(k') &= \beta V'(k'). \end{aligned}$$

Let $\tilde{k}>k$. Then $f(\tilde{k})-k'>f(k)-k'$. Strict concavity of u implies that $u'\left[f(\tilde{k})-k'\right]<u'\left[f(k)-k'\right]$. Hence we have that

$$\tilde{k}>k\Rightarrow LHS(\tilde{k},k')<LHS\,(k,k')\,.$$

As a consequence, the $RHS\,(k')$ must decrease to satisfy the first-order condition. Since $V'(\cdot)$ is decreasing, this will occur only if k' increases. This shows that $\tilde{k}>k\Rightarrow g(\tilde{k})>g(k)$.

The above result can also be viewed as an application of the implicit function theorem. Define

$$H(k,k')\equiv u'\left[f(k)-k'\right]-\beta V'(k')=0.$$

Then

$$\begin{aligned} \frac{\partial k'}{\partial k} &= -\frac{\dfrac{\partial H(k,k')}{\partial k}}{\dfrac{\partial H(k,k')}{\partial k'}} \\ &= -\frac{u''\left[f(k)-k'\right]f'(k)}{-u''\left[f(k)-k'\right]-\beta V''(k')} \\ &= \frac{\overbrace{u''\left[f(k)-k'\right]}^{(-)}\overbrace{f'(k)}^{(+)}}{\underbrace{u''\left[f(k)-k'\right]}_{(-)}+\underbrace{\beta V''(k')}_{(-)}}>0, \end{aligned}$$

where the sign follows from the fact that since u and V are strictly concave and f is strictly increasing, both the numerator and the denominator of this expression have negative signs. This derivation is heuristic since we have assumed here that V is twice continuously differentiable. It turns out that there is a theorem telling us that (under some side conditions that we will not state here) V will indeed be twice continuously differentiable, given that u and f are both twice differentiable, but it is beyond the scope of the present analysis to discuss this theorem in greater detail. ■

The economic intuition behind g being increasing is simple. There is an underlying presumption of normal goods behind our assumptions: strict concavity and additivity of the different consumption goods (over time) amounts to assuming that the different goods are normal goods. Specifically, consumption in the future is a normal good. Therefore, a larger initial wealth commands larger savings.

(vi) $c(k) \equiv f(k) - g(k)$ is strictly increasing and, hence, the marginal propensities to consume and to save out of income are both strictly between 0 and 1.

Proof. In line with the previous proof, write the first-order condition as

$$u'\left[c\right] = \beta V'(f(k) - c).$$

$V'(\cdot)$ is decreasing, since $V(\cdot)$ is strictly concave due to the assumptions on u and f. Define

$$\begin{aligned} LHS\left(c\right) &= u'\left[c\right] \\ RHS\left(k, c\right) &= \beta V'(f(k) - c). \end{aligned}$$

Let $\tilde{k} > k$. Then $f(\tilde{k}) - c > f(k) - c$. Strict concavity of V implies that $V'\left[f(\tilde{k}) - c\right] < V'\left[f(k) - c\right]$. Hence we have that

$$\tilde{k} > k \Rightarrow RHS(\tilde{k}, c) < RHS\left(k, c\right).$$

As a consequence, c must change to $\tilde{c}$ in response to a change from k to $\tilde{k}$ so as to counteract the decrease in RHS. This is not possible unless $\tilde{c} > c$. So, by means of contradiction, suppose that $\tilde{c} < c$. Then, since u is strictly concave, LHS would rise and, since V is also strictly concave, RHS would decrease further, increasing rather than decreasing the gap between the two expressions. It follows that $c(k)$ must be (globally) increasing. Together with the previous fact, we conclude that an increase in k would both increase consumption and investment. Put in terms of an increase in output $f(k)$, an increase in output would lead to a less than one-for-one increase both in consumption and investment. ■

(vii) $g(k^*) = k^*$, where k^* solves $\beta f'(k^*) = 1$.

The functional Euler equation reads

$$u'(f(k) - g(k)) = \beta u'(f(g(k)) - g(g(k)))f'(g(k)).$$

It is straightforward to verify that the guess solves this equation at k^*. However, is this the only value for $g(k^*)$ that solves the equation at $k = k^*$? Suppose, by means of contradiction, that $g(k^*) > k^*$ or that $g(k^*) < k^*$. In the former case, then, relative to the case of our prime candidate $g(k^*) = k^*$, the left-hand side of the equation rises, since u is strictly concave. Moreover, the right-hand side falls, for two reasons: $f(y) - g(y)$ is increasing, as shown above, so strict concavity of u implies that $u'(f(g(k^*)) - g(g(k^*))) < u'(f(k^*) - g(k^*))$; and strict concavity of f' implies that $f'(g(k^*)) < f'(k^*)$. Using parallel reasoning, the latter case, i.e., $g(k^*) < k^*$, leads to an increase in the right-hand side and a decrease in the left-hand side. Thus, neither case allows the Euler equation to be satisfied at k^*. Hence, since $g(k^*)$ exists, and since it must satisfy the Euler equation, it must equal k^*.

(viii) $g(k)$, if differentiable, has slope less than one at k^*.

Differentiation and evaluation of the functional Euler equation at k^* delivers

$$u''(f' - g') = \beta \left(u''(f' - g')g'f' + u'f''g'\right)$$

from which follows, since $\beta f' = 1$, that

$$(f' - g')(1 - g') = \frac{u'f''}{u''f'}g'.$$

This means, since we have shown above that g' and $f' - g'$ are above zero, that $g' < 1$ must hold.

9.2.2 Global convergence

In order to discuss convergence to a unique steady state, we will first make somewhat heuristic use of the properties of g established above to suggest that there is global convergence to k^*. We will then offer a full, and entirely independent, proof of global convergence.

We know that $g(k)$ has to start out at 0, be continuous and increasing, and satisfy $g(\bar{k}) \leq \bar{k}$ (the latter, in fact, with inequality, because $u'(0) = \infty$). Now let us consider some different possibilities for the decision rule. Figure 9.2 shows three decision rules which all share the mentioned properties.

Line 1 has three different solutions to the steady-state condition $k' = k$, line 2 has only one steady state and line 3 has no positive steady state.

Line 1 can be ruled out because if there is a steady state, it is unique, due to strict concavity of u. Similarly, this argument rules out any decision rule with more than one positive crossing of the 45^o line.

Line 3, with no positive steady state, can be ruled out since it contradicts property (vii): there is a steady state at k^*. Similarly, a decision rule that starts, and remains, above the 45^o line over the entire domain $[0, \bar{k}]$ would not be possible (though we also know that $g(\bar{k})$ must be below $\bar{k}$ to ensure positive consumption).

Has any possibility been ruled out, or is the only remaining possibility line 2? In fact, there is one more possibility, namely, that $g(k)$ starts out below the 45^o line, increases toward

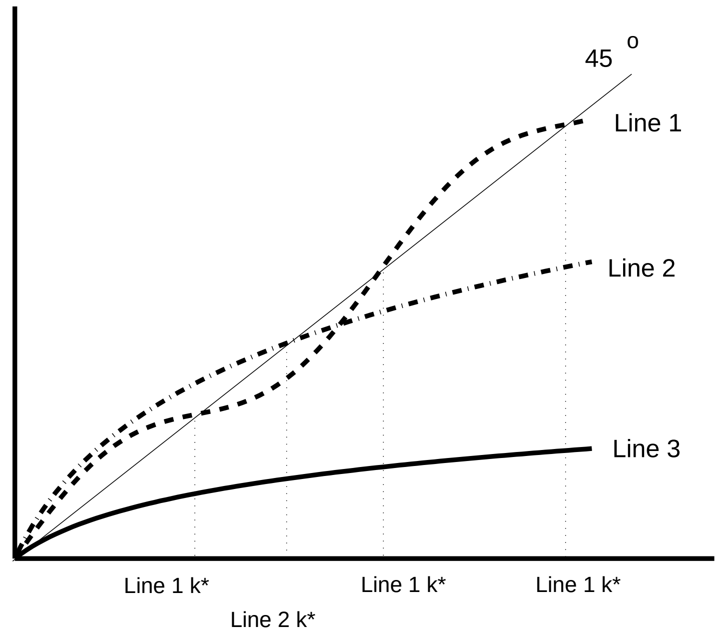

Figure 9.2: Different decision rule candidates

the steady state k^*, then "touches" k^* and, for $k > k^*$, again falls below the 45^o line. Is this possibility ruled out by the facts above? It is: it would require that $g(k)$, at least in a left neighborhood of k^*, increases more than one-for-one. But we have established above, with result (viii), that its slope must be smaller than one at the steady state.

Having ruled out all alternatives, line 2 is clearly above the 45^o line to the left of k^*, and below to the right. This implies that the model dynamics exhibit global convergence.

The convergence will not occur in finite time. For it to occur in that manner, the decision rule would have to be flat at the steady state point. This, however, cannot be since we have established that $g(k)$ is strictly increasing (Property 2).

Let us now, formally, turn to a significantly more elegant proof of global convergence. From the fact that V is strictly concave, we know that

$$\left(V'(k) - V'(g(k))\right)\left(k - g(k)\right) \leq 0$$

with strict equality whenever $k \neq g(k)$. Since $V'(k) = u'(f(k) - g(k))f'(k)$ from the envelope theorem and, from the first-order condition, $\beta V'(g(k)) = u'(f(k) - g(k))$, we obtain

$$\left(\beta f'(k) - 1\right)\left(k - g(k)\right) \leq 0$$

using the fact that $u'(c) \geq 0$. Thus, it follows directly that $g(k)$ exceeds k whenever k is below k^* (since f is strictly concave), and vice versa.

9.2.3 Dynamics: the speed of convergence

What can we say about the time it takes to reach the steady state? The speed of global convergence will depend on the shape of $g(k)$, as Figure 9.3 shows.

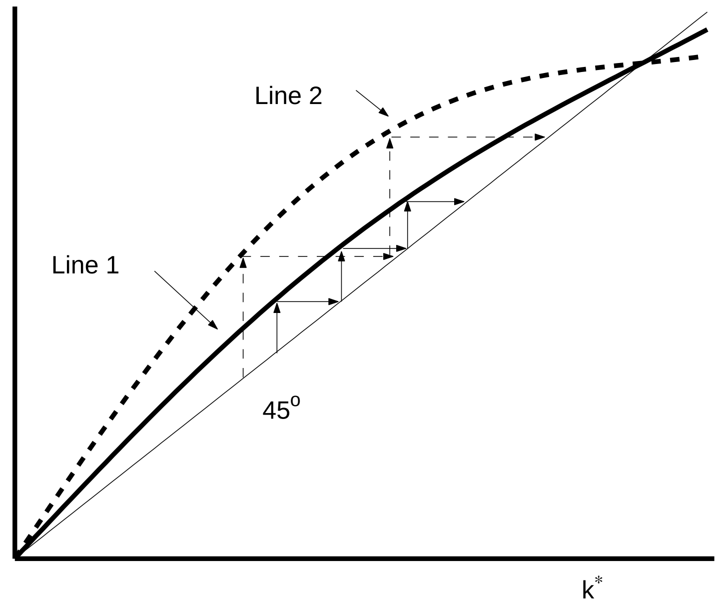

Figure 9.3: Different speeds of convergence

Capital will approach the steady state level more rapidly (i.e., in "a smaller number of steps") along trajectory number 2, where it will have a faster *speed of convergence*. There is no simple way to summarize, in a quantitative way, the speed of convergence for a general decision rule. However, for a limited class of decision rules - the linear (or affine) rules - it can be measured simply by looking at the slope. This is an important case, for it can be used locally to approximate the speed of convergence around the steady state k^*.

The argument for this is simple: the accumulation path will spend infinite time arbitrarily close to the steady state, and in a very small region a continuous function can be arbitrarily well approximated by a linear function, using the first-order Taylor expansion of the function. That is, for any capital accumulation path, we will be able to approximate the speed of convergence arbitrarily well as time passes. If the starting point is far from the steady state, we will make mistakes that might be large initially, but these mistakes will become smaller and smaller and eventually become unimportant. Moreover, if one uses parameter values that are, in some sense, realistic, it turns out that the resulting decision rule will be quite close to a linear one.

In this section, we will state a general theorem with properties for dynamic systems of a general size. To be more precise, we will be much more general than the one-sector growth model. With the methods we describe here it is actually possible to obtain the key information about local dynamics for any dynamic system. The global convergence theorem, in contrast, applies only for the one-sector growth model.

The first-order Taylor series expansion of the decision rule gives

$$\begin{aligned} k' &= g(k) \approx \underbrace{g(k^*)}_{k^*} + g'(k^*)\,(k-k^*) \\ \underbrace{k'-k^*}_{\text{Next period's gap}} &= g'(k^*)\,\underbrace{(k-k^*)}_{\text{Current gap}}. \end{aligned}$$

This shows that we may interpret $g'(k^*)$ as a measure of the rate of convergence (or rather, its inverse). If $g'(k^*)$ is very close to zero, convergence is fast and the gap decreases significantly each period.

Linearization for a general dynamic system

The task is now to find $g'(k^*)$ by *linearization*. We will use the Euler equation and linearize it. This will lead to a difference equation in k_t. One of the solutions to this difference equation will be the one we are looking for. Two natural questions arise: 1) How many convergent solutions are there around k^*? 2) For the convergent solutions, is it valid to analyze a linear difference equation as a proxy for their convergence speed properties? The first of these questions is the key to the general characterization of dynamics. The second question is a mathematical one and related to the approximation precision.

Both questions are addressed by the following theorem, which applies to a general dynamic system (i.e., not only those coming from economic models):

Theorem 9.1 *Let $x_t \in \Re^n$. Given $x_{t+1} = h(x_t)$ with a stationary point $\bar{x} : \bar{x} = h(\bar{x})$. If*

1. *h is continuously differentiable with Jacobian $H(\bar{x})$ around $\bar{x}$ and*
2. *$I - H(\bar{x})$ is non-singular,*

then there is a set of initial conditions x_0, of dimension equal to the number of eigenvalues of $H(\bar{x})$ that are less than 1 in absolute value, for which $x_t \to \bar{x}$.

The idea behind the proof, and the usefulness of the result, relies on the idea is that, close enough to the stationary point, the nonlinear dynamic system behaves like its linear(ized) counterpart. Letting $H(\bar{x}) \equiv H$, the linear counterpart would read $x_{t+1} - \bar{x} = H(x_t - \bar{x})$. Assuming that H can be diagonalized with distinct eigenvalues collected in the diagonal matrix $\Lambda(\bar{x})$, so that $H = B^{-1}\Lambda B$ with B being a matrix of eigenvectors, the linear system can be written

$$B(x_{t+1} - \bar{x}) = \Lambda B(x_t - \bar{x})$$

and, hence,

$$B(x_t - \bar{x}) = \Lambda^t B(x_0 - \bar{x}).$$

Here, with distinct eigenvalues it is straightforward to see that whether $B(x_t - \bar{x})$ will go to zero (and hence x_t converge to $\bar{x}$) will depend on the size of the eigenvalues and on the initial vector x_0.

We will describe how to use these results with a few examples.

Example 9.2 ($n = 1$) *There is only one eigenvalue: $\lambda = h'(\bar{x})$*

1. $|\lambda| \geq 1 \Rightarrow$ *no initial condition leads to x_t converging to $\bar{x}$.*

   In this case, only for $x_0 = \bar{x}$ will the system stay in $\bar{x}$.

2. $|\lambda| < 1 \Rightarrow x_t \to \bar{x}$ *for any value of x_0.*

Example 9.3 ($n = 2$) *There are two eigenvalues λ_1 and λ_2.*

1. $|\lambda_1|, |\lambda_2| \geq 1 \quad \Rightarrow$ *No initial condition x_0 leads to convergence.*

2. $|\lambda_1| < 1$, $|\lambda_2| \geq 1 \Rightarrow$ *Dimension of x_0's leading to convergence is 1. This is called "saddle path stability".*

3. $|\lambda_1|, |\lambda_2| < 1 \quad \Rightarrow$ *Dimension of x_0's leading to convergence is 2. $x_t \to \bar{x}$ for any value of x_0.*

The examples describe how a general dynamic system behaves. It does not yet, however, quite settle the issue of convergence. In particular, the set of initial conditions leading to convergence must be given an economic meaning. Is any initial condition possible in a given economic model? Typically no: for example, the initial capital stock in an economy may be given, and thus we have to restrict the set of initial conditions to those respecting the initial capital stock.

We will show below that an economic model has dynamics that can be reduced to a vector difference equation of the form of the one described in the above theorem. In this description, the vector will have a subset of true state variables (e.g. capital) while the remainder of the vector consists of various control, or other, variables that are there in order that the system can be put into first-order form.

More formally, let the number of eigenvalues less than 1 in absolute value be denoted by m. This is the dimension of the set of initial x_0's leading to $\bar{x}$. We may interpret m as the *degrees of freedom*. Let the number of (distinct) economic restrictions on initial conditions be denoted by $\hat{m}$. These are the restrictions emanating from physical (and perhaps other) conditions in our economic model. Notice that an interpretation of this is that we have $\hat{m}$ equations and m unknowns. Then the issue of convergence boils down to the following cases.

1. $m = \hat{m} \Rightarrow$ there is a *unique* convergent solution to the difference equation system.

2. $m < \hat{m} \Rightarrow$ No convergent solution obtains.

3. $m > \hat{m} \Rightarrow$ There is "indeterminacy", i.e., there are many convergent solutions (how many? dim $= \hat{m} - m$).

Solving for the speed of convergence

We now describe in detail how the linearization procedure works. The example comes from the one-sector growth model, but the general outline is the same for all economic models.

1. Derive the Euler equation: $F(k_t, k_{t+1}, k_{t+2}) = 0$

   $$u'[f(k_t) - k_{t+1}] - \beta u'[f(k_{t+1}) - k_{t+2}] f'(k_{t+1}) = 0.$$

   Clearly, k^* is a steady state $\Leftrightarrow F(k^*, k^*, k^*) = 0$.

2. Linearize the Euler equation: Define $\hat{k}_t = k_t - k^*$ and using first-order Taylor approximation derive a_0, a_1, and a_2 such that

   $$a_2 \hat{k}_{t+2} + a_1 \hat{k}_{t+1} + a_0 \hat{k}_t = 0.$$

3. Write the Euler equation as a first-order system: A difference equation of any order can be written as a first order difference equation by using vector notation: Define $\mathrm{x}_t = \begin{pmatrix} \hat{k}_{t+1} \\ \hat{k}_t \end{pmatrix}$ and then

   $$x_{t+1} = H x_t.$$

4. Find the solution to the first-order system: Find the unknowns in

   $$x_t = c_1 \lambda_1^t v_1 + c_2 \lambda_2^t v_2, \tag{9.1}$$

   where c_1 and c_2 are constants to be determined, λ_1 and λ_2 are (distinct) eigenvalues of H, and v_1 and v_2 are eigenvectors associated with these eigenvalues.

5. Determine the constants: Use the information about state variables and initial conditions to find c_1 and c_2. In this case, x consists of one state variable and one lagged state variable, the latter used only for the reformulation of the dynamic system. Therefore, we have one initial condition for the system, given by k_0; this amounts to one restriction on the two constants. The set of initial conditions for x_0 in our economic model has therefore been reduced to one dimension. Finally, we are looking for convergent solutions. If one of the two eigenvalues is greater than one in absolute value, this means that we need to set the corresponding constant to zero. Consequently, since not only k_0 but also k_1 are now determined (i.e., both elements of x_0), and our system is fully determined: all future values of k (or x) can be obtained.

   If both eigenvalues are larger than one, the dynamics will not have convergence to the steady state: only if the system starts at the steady state will it remain there.

   If both eigenvalues are less than one, we have no way of pinning down the remaining constant, and the set of converging paths will remain of one dimension. Such indeterminacy - effectively an infinite number of solutions to the system - will not occur in our social planning problem, because (under strict concavity) it is guaranteed that the set of solutions is a singleton. However, in equilibrium systems that are not derived from a planning problem (perhaps because the equilibrium is not Pareto optimal, as we shall see below), it is possible to end up with indeterminacy.

The typical outcome in our one-sector growth model is $0 < \lambda_1 < 1$ and $\lambda_2 > 1$, which implies $m = 1$ (saddle path stability). Then the convergent solution has $c_2 = 0$. In other words, the economics of our model dictate that the number of restrictions we have on the initial conditions is one, namely the (given) initial level of capital, k_0, i.e. $\hat{m} = 1$. Therefore, $m = \hat{m}$, so there is a unique convergent path for each k_0 (close to k^*).

Then c_1 is determined by setting $c_2 = 0$ (so that the path is convergent) and solving equation (9.1) for the value of c_1 such that if $t = 0$, then k_t is equal to the given level of initial capital, k_0.

We now implement these steps in detail for a one-sector optimal growth model.

First, we need to solve for H. Let us go back to

$$u'\left[f(k_t) - k_{t+1}\right] - \beta u'\left[f(k_{t+1}) - k_{t+2}\right] f'(k_{t+1}) = 0.$$

In order to linearize it, we take derivatives of this expression with respect to k_t, k_{t+1} and k_{t+2}, and evaluate them at k^*. We obtain

$$\beta u''(c^*)f'(k^*)\hat{k}_{t+2} - \left[u''(c^*) + \beta u''(c^*)\left[f'(k^*)\right]^2 + \beta u'(c^*)f''(k^*)\right]\hat{k}_{t+1} +$$

$$+ u''(c^*)f'(k^*)\hat{k}_t = 0.$$

Using the steady-state fact that $\beta f'(k^*) = 1$, we simplify this expression to

$$u''(c^*)\hat{k}_{t+2} - \left[u''(c^*) + \beta^{-1}u''(c^*) + u'(c^*)\left[f'(k^*)\right]^{-1} f''(k^*)\right]\hat{k}_{t+1} + \beta^{-1}u''(c^*)\hat{k}_t = 0.$$

Dividing through by $u''(c^*)$, we arrive at

$$\hat{k}_{t+2} - \left[1 + \frac{1}{\beta} + \frac{u'(c^*)}{u''(c^*)}\frac{f''(k^*)}{f'(k^*)}\right]\hat{k}_{t+1} + \frac{1}{\beta}\hat{k}_t = 0.$$

Then

$$\begin{pmatrix} \hat{k}_{t+2} \\ \hat{k}_{t+1} \end{pmatrix} = H \begin{pmatrix} \hat{k}_{t+1} \\ \hat{k}_t \end{pmatrix}$$

with

$$H = \begin{pmatrix} 1 + \frac{1}{\beta} + \frac{u'(c^*)}{u''(c^*)}\frac{f''(k^*)}{f'(k^*)} & -\frac{1}{\beta} \\ 1 & 0 \end{pmatrix}.$$

This is a second-order difference equation. Notice that the second row of H delivers $\hat{k}_{t+1} = \hat{k}_{t+1}$, so the vector representation of the system is correct. Now we need to look for the eigenvalues of H, from the characteristic polynomial given by

$$|H - \lambda I| = 0.$$

As an interlude before solving for the eigenvalues, let us now motivate the general solution to the linear system above with an explicit derivation from basic principles. Using spectral decomposition, we can decompose H as follows:

$$H = V \Lambda V^{-1} \Rightarrow \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

where λ_1 and λ_2 are eigenvalues of H and V is a matrix of eigenvectors of H. Recall that

$$x_{t+1} = H x_t.$$

A change of variables will help us get the solution to this system. First premultiply both sides by V^{-1}:

$$\begin{aligned} V^{-1} x_{t+1} &= V^{-1} H x_t \\ &= V^{-1} V \Lambda V^{-1} x_t \\ &= \Lambda V^{-1} x_t. \end{aligned}$$

Let $z_t \equiv V^{-1} x_t$ and $z_{t+1} \equiv V^{-1} x_{t+1}$. Then, since Λ is a diagonal matrix

$$\begin{aligned} z_{t+1} &= \Lambda z_t \\ z_t &= \Lambda^t z_0 \\ z_{1t} &= c_1 \lambda_1^t = z_{10} \lambda_1^t \\ z_{2t} &= z_{20} \lambda_2^t. \end{aligned}$$

We can go back to x_t by premultiplying z_t by V:

$$\begin{aligned} x_t &= V z_t \\ &= V \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} \\ &= c_1 \lambda_1^t \begin{pmatrix} V_{11} \\ V_{21} \end{pmatrix} + c_2 \lambda_2^t \begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} \\ &= \begin{pmatrix} \hat{k}_{t+1} \\ \hat{k}_t \end{pmatrix}. \end{aligned}$$

The solution, therefore must be of the form

$$\hat{k}_t = \widehat{c}_1 \lambda_1^t + \widehat{c}_2 \lambda_2^t,$$

where $\widehat{c}_1$ and $\widehat{c}_2$ are to be determined from initial conditions and values of λ_1 and λ_2.

Let us now go back to our example. To find the eigenvalues in our specific setting, we use $|H - \lambda I| = 0$ to obtain

$$\begin{vmatrix} 1 + \frac{1}{\beta} + \frac{u'}{u''}\frac{f''}{f'} - \lambda & -\frac{1}{\beta} \\ 1 & -\lambda \end{vmatrix} = 0$$

$$\Rightarrow \lambda^2 - \left[1 + \frac{1}{\beta} + \frac{u'}{u''}\frac{f''}{f'}\right]\lambda + \frac{1}{\beta} = 0, \tag{9.2}$$

where u', u'', f', f'' denote the corresponding derivatives evaluated at k^*. Let

$$F(\lambda) \equiv \lambda^2 - \left[1 + \frac{1}{\beta} + \frac{u'}{u''}\frac{f''}{f'}\right]\lambda + \frac{1}{\beta}.$$

This is a continuous function of λ, and

$$\begin{aligned} F(0) &= \frac{1}{\beta} > 0 \\ F(1) &= -\frac{u'}{u''}\frac{f''}{f'} < 0. \end{aligned}$$

Therefore, the mean value theorem implies that $\exists \lambda_1 \in (0,1) : F(\lambda_1) = 0$. That is, one of the eigenvalues is positive and smaller than one. Since $\lim_{\lambda \to \infty} F(\lambda) = +\infty > 0$, the other eigenvalue ($\lambda_2$) must also be positive and larger than 1.

We see that a convergent solution to the system requires $c_2 = 0$. The remaining constant, c_1, will be determined from

$$\begin{aligned} \hat{k}_t &= \widehat{c}_1 \lambda_1^t \\ \hat{k}_0 &\equiv k_0 - k^* \\ &\Rightarrow \widehat{c}_1 = k_0 - k^*. \end{aligned}$$

The solution, therefore, is

$$k_t = k^* + \lambda_1^t (k_0 - k^*).$$

Recall that

$$k_{t+1} - k^* = g'(k^*)(k_t - k^*).$$

Analogously, in the linearized system,

$$k_{t+1} - k^* = \lambda_1 (k_t - k^*).$$

It can thus be seen that the eigenvalue λ_1 has a particular meaning: it measures the (inverse of the) rate of convergence to the steady state.

As a different illustration, suppose we were looking at the larger system

$$\begin{aligned} k_t &= c_1\lambda_1^t + c_2\lambda_2^t + c_3\lambda_3^t + c_4\lambda_4^t, \\ & k_0 \text{ given.} \end{aligned}$$

That is, some economic model with a single state variable leads to a third-order difference equation. If only one eigenvalue λ_1 has $|\lambda_1| < 1$, then there is a unique convergent path leading to the steady state. This means that c_2, c_3, c_4, will need to be equal to zero (choosing the subscript 1 to denote the eigenvalue smaller than 1 in absolute value is arbitrary, of course).

In contrast, if there were, for example, two eigenvalues λ_1, λ_2 with $|\lambda_1|, |\lambda_2| < 1$, then we would have $m = 2$ (two "degrees of freedom"). But there is only one economic restriction, namely k_0 given. That is, $\hat{m} = 1 < m$. Then there would be many convergent paths satisfying the sole economic restriction on initial conditions and the system would be indeterminate.

Alternative solution to the speed of convergence

There is another way to solve for the speed of convergence. It is related to the argument that we have local convergence around k^* if the slope of the $g(k)$ schedule satisfies $g'(k^*) \in (-1, 1)$.

The starting point is the functional Euler equation:

$$u'[f(k) - g(k)] = \beta u'[f(g(k)) - g(g(k))]f'(g(k)), \ \forall k.$$

Differentiating with respect to k yields

$$\begin{aligned} u''[f(k) - g(k)][f'(k) - g'(k)] &= \beta u''[f(g(k)) - g(g(k))][f'(g(k))g'(k) - g'(g(k))g'(k)] \times \\ &\quad \times f'(g(k)) + \beta u'[f(g(k)) - g(g(k))]f''(g(k))g'(k), \ \forall k. \end{aligned}$$

Evaluating at the steady state and noting that $g(k^*) = k^*$, we get

$$u''(c^*)[f'(k^*) + g'(k^*)] = \beta u''(c^*)[f'(k^*)g'(k^*) - (g'(k^*))^2]f'(k^*) + \beta u'(c^*)f''(k^*)g'(k^*).$$

This equation is a quadratic equation in $g'(k^*)$. Reshuffling the terms and noting that $\beta f'(k^*) = 1$, we are lead back to equation (9.2) from before with the difference that we have now $g'(k^*)$ instead of λ. Using the same assumptions on $u(\cdot)$ and $f(\cdot)$, we can easily prove that for one of the solutions $g_1'(k^*) \in (-1, 1)$. The final step is the construction of $g(k)$ using a linear approximation around k^*.

9.3 Growth theory II: exogenous technological change

In this section we will study the basic framework to model output growth by introducing an exogenous change in the production technology that takes place over time. Mathematically, this is just a simple modification of the standard neoclassical growth model that we have seen before.

We will separate the issue of growth into two components. One is a technological component: is growth feasible with the assumed production technology? The second one is the decision making aspect involved: will a central planner, or the decentralized economy, choose a growing path? Which types of utility function allow for what we will call a "balanced growth path"?

This section is split into three subsections. The first and second ones address the technological and decision making issues, respectively. In the third one, we will study a transformation to the exogenous growth model that will help us in the analysis.

9.3.1 Exogenous long-run growth

Balanced growth under labor-augmenting technological change

Given the assumptions regarding the production technology on the one hand, and regarding the source of technological progress on the other, we want to analyze whether the standard neoclassical growth model is really consistent with sustained output growth. From the point of view of the production side, is sustainable output growth feasible?

The standard case is that of labor-augmenting technological change (à la Solow). The resource constraint in the economy is:

$$c_t + i_t = F_t(K_t, \underbrace{n_t}_{\text{hours}}) \quad = F(K_t, \underbrace{\gamma^t n_t}_{\text{"efficiency units"}}),$$

where F represents a constant returns to scale production technology and $\gamma > 1$. The capital accumulation law is

$$k_{t+1} = (1 - \delta)\, k_t + i_t.$$

Given the constant returns to scale assumption on F, sustained growth is then possible. Let us analyze this setup in detail.

Our object of study is what is called *balanced growth*: all economic variables grow at *constant* rates (that could vary from one variable to another). In this case, this would imply that for all t, the value of each variable in the model is given by:

$$\left.\begin{array}{l} y_t = y_0 g_y^t \\ c_t = c_0 g_c^t \\ k_t = k_0 g_k^t \\ i_t = i_0 g_i^t \\ n_t = n_0 g_n^t. \end{array}\right\} \begin{array}{l}\text{balanced growth path -}\\ \text{all variables grow at constant}\\ \text{(but possibly different) rates.}\end{array}$$

In a model with growth, this is the analogue of a steady state.

Our task is to find the growth rate for each variable in a balanced growth path, and check whether such a path is consistent. We begin by guessing one of the growth rates, as follows. From the capital accumulation law

$$k_{t+1} = (1 - \delta)\, k_t + i_t.$$

If both i_t and k_t are to grow at a constant rate, it must be the case that they both grow at the same rate, i.e., $g_k = g_i$. By the same type of reasoning, from the resource constraint

$$c_t + i_t = F_t(k_t,\, n_t) = F(k_t,\, \gamma^t n_t) \equiv y_t$$

we must have that $g_y = g_c = g_i$.

Next, using the fact that F represents a constant-returns-to-scale technology (and hence it is homogenous of degree one), we have that

$$F(k_t,\, \gamma^t n_t) = \gamma^t n_t F\left(\frac{k_t}{\gamma^t n_t},\, 1\right)$$

$$\Rightarrow \frac{y_t}{\gamma^t n_t} = F\left(\frac{k_t}{\gamma^t n_t},\, 1\right).$$

Since we have postulated that k_t and y_t grow at a constant rate, we must have that

$$\frac{k_t}{\gamma^t n_t} = \text{constant}.$$

In addition, since the time endowment is bounded, actual hours can not grow beyond a certain upper limit (usually normalized to 1); hence $g_n = 1$ must hold.

This results in $g_k = \gamma$, and all other variables also grow at rate γ. Hence, it is possible to obtain constant growth for all variables: a balanced growth path is technologically feasible.

The nature of technological change

From the analysis in the previous section, it seems natural to ask whether the assumption that the technological change is labor-augmenting is relevant or not. First, what other kinds of technological change can we think of? On a more general level than that described above, ignoring labor input for a moment, an intertemporal production possibility set (through the accumulation of capital) involves some function of consumption outputs at different points in time, such as

$$G(c_0, c_1, \dots) = 0,$$

and technological change—or "productivity growth"—implies that G is asymmetric with respect to its different arguments, in effect tilting the production possibility set towards consumption in the future. Such tilting can take many forms, and a general discussion of how data can allow us to distinguish different such forms is beyond the scope of the discussion here. In practical modeling, one typically encounters parameterizations of technological change of the sort described above: a constantly shifting factor, say, multiplying one production input. The purpose of the ensuing discussion is to describe some commonly used such forms of technological change and the feasibility of balanced growth in these cases.

Let us first write the economy's resource constraint, with all the technology shift factors that are most commonly emphasized in the literature (maintaining $k_{t+1} = (1-\delta)k_t + i_t$):

$$c_t + \gamma_{it} i_t = \gamma_{zt} F\left(\gamma_{kt} k_t,\, \gamma_{nt} n_t\right).$$

The associated nomenclature is as follows.

- γ_{nt}: Labor-augmenting technological change: a rise in this parameter from one period to the next raises the effective value of the total labor input.

- γ_{kt}: Capital-augmenting technological change: a rise in this parameter raises the effective value of any given capital stock.

- γ_{zt}: Neutral (or Hicks-neutral) technological change: a rise in this parameter raises output proportionally, for given inputs.

- γ_{it}: Investment-specific technological change: a fall in this parameter makes it cheaper to produce capital; thus, it makes additions to the capital stock easier to obtain (as opposed to a change in $\gamma_{k,t+1}$ which raises the value of the entire stock of capital).

Given the assumption of constant returns to scale, one can subsume γ_{zt} in γ_{kt} and γ_{nt}, so we will set $\gamma_{zt} = 1$ for all t from here and on. Also, given that both investment-specific and capital-augmenting technological change operate through the capital stock, they are closely related: an increase in γ_{kt} would appear very similar to appropriately increasing the prior sequence of γ_{is}, for given values of i_s, $s < t$, although changes in the latter will influence the stock of capital at dates after t. Formally, we can define $\hat{i}_t \equiv i_t\gamma_{it}$ as investment *in consumption units*, and similarly $\hat{k}_t \equiv k_t\gamma_{i,t-1}$ and write the economy as

$$c_t + \hat{i}_t = F\left(\hat{\gamma}_{kt}\hat{k}_t,\, \gamma_{nt}n_t\right),$$

with the new capital accumulation equation

$$\hat{k}_{t+1} = \hat{i}_t + (1-\delta_t)\hat{k}_t,$$

where $\hat{\gamma}_{kt} \equiv \gamma_{kt}/\gamma_{i,t-1}$ and $\delta_t \equiv \delta(\gamma_{it}/\gamma_{i,t-1}) + 1 - (\gamma_{it}/\gamma_{i,t-1})$, which is in $(0,\delta)$ whenever $\gamma_{it} < \gamma_{i,t-1}$ and $\delta \in (0,1)$. Thus, this formulation makes clear that we can think of investment-specific technological change in terms of depreciating the existing capital stock, measured in consumption units, at a higher rate than the physical wear-and-tear rate δ; in other respects, capital-augmenting and investment-specific technological change are identical, since they both effectively enhance the stock of capital, measured in consumption units, and thus improve the consumption possibilities over time (provided that $\gamma_{kt} > 1 > \gamma_{it}$).

Now suppose that we consider balanced growth paths, with restricted attention to technology factors growing at constant rates: $\gamma_{it} = \gamma_i^{-t}$, $\gamma_{kt} = \gamma_k^t$, and $\gamma_{nt} = \gamma_n^t$, with γ_i, γ_k, and γ_n all greater than or equal to 1. Can we have balanced growth in this economy when one or several of these growth factors are strictly larger than one? We have the following result.

Theorem 9.4 *For exact balanced growth, $\gamma_i = \gamma_k = 1$ need to hold (thus, only allowing $\gamma_n > 1$), unless F is a Cobb-Douglas function.*

Proof. In one of the directions, the proof requires an argument involving partial differential equations which we shall not develop here. However, we will show that if F is a Cobb-Douglas function then any of the γs can be larger than 1, without invalidating a balanced growth path as a solution.

If F is a Cobb-Douglas function, the resource constraint reads:

$$c_t + \gamma_i^{-t} i_t = \left(\gamma_k^t k_t\right)^\alpha \left(\gamma_n^t n_t\right)^{1-\alpha}. \tag{9.3}$$

Notice that we can define

$$\widehat{\gamma}_n \equiv \gamma_k^{\frac{\alpha}{1-\alpha}} \gamma_n$$

so that we can rewrite the production function:

$$\left(\gamma_k^t k_t\right)^\alpha \left(\gamma_n^t n_t\right)^{1-\alpha} = k_t^\alpha \left(\widehat{\gamma}_n^t n_t\right)^{1-\alpha}. \tag{9.4}$$

We will use this formulation later.

Now consider the capital accumulation equation:

$$k_{t+1} = (1-\delta) k_t + i_t.$$

Dividing through by γ_i^t, we obtain

$$\frac{k_{t+1}}{\gamma_i^{t+1}} \gamma_i = (1-\delta) \frac{k_t}{\gamma_i^t} + \frac{i_t}{\gamma_i^t}.$$

We can define

$$\widetilde{k}_t \equiv \frac{k_t}{\gamma_i^t}, \; \widetilde{i}_t \equiv \frac{i_t}{\gamma_i^t}$$

and, replacing $\widetilde{k}_t$ in (9.3), we obtain:

$$\begin{aligned} c_t + \widetilde{i}_t &= \left(\gamma_k^t \gamma_i^t \widetilde{k}_t\right)^\alpha \left(\gamma_n^t n_t\right)^{1-\alpha} \\ \widetilde{k}_{t+1} \gamma_i &= (1-\delta) \widetilde{k}_t + \widetilde{i}_t. \end{aligned}$$

The model has been transformed into an equivalent system in which $\widetilde{k}_{t+1}$, instead of k_{t+1}, is the object of choice (more on this below). Notice that since F is Cobb-Douglas, γs multiplying $\widetilde{k}_t$ can in fact be written as labor-augmenting technological growth factors (see (9.4)). Performing the transformation, the rate of growth in labor efficiency units is

$$\gamma_n \gamma_k^{\frac{\alpha}{1-\alpha}} \gamma_i^{\frac{\alpha}{1-\alpha}},$$

and we have seen above that this is also the growth rate of output and consumption. ■

Convergence in the neoclassical growth model

Consider first Solow's model without population growth, i.e., let the savings rate be exogenously given by s. In transformed form, so that $\hat{y}_t \equiv y_t/\gamma^t$, where γ is the growth rate of labor-augmenting technology, we obtain

$$\hat{y}_{t+1} = \gamma^{-(t+1)} F(sF(k_t, \gamma^t) + (1-\delta)k_t, \gamma^{t+1}) = F(\gamma^{-1} sF(\hat{k}_t, 1) + (1-\delta)\hat{k}_t, 1)$$

so that

$$\hat{y}_{t+1} = f(\gamma^{-1} s\hat{y}_t + \gamma^{-1}(1-\delta) f^{-1}(\hat{y}_t)).$$

Assuming that F is Cobb-Douglas, with a capital share of α, we can obtain a closed-form solution for $d\hat{y}_{t+1}/d\hat{y}_t$ evaluated at steady state. Taking derivatives we obtain

$$\frac{d\hat{y}_{t+1}}{d\hat{y}_t} = f' \frac{s + (1-\delta)\frac{1}{f'}}{\gamma} = \frac{s\alpha(\hat{y}/\hat{k}) + 1 - \delta}{\gamma} = \frac{\alpha(\gamma - 1 + \delta) + 1 - \delta}{\gamma} = \alpha + (1-\alpha)\frac{1-\delta}{\gamma},$$

where we also used the balanced-growth relation $\gamma\hat{k} = s\hat{y} + (1-\delta)\hat{k}$. Notice that we could alternatively have derived this as $d\hat{k}_{t+1}/d\hat{k}_t$.

The growth regression reported above was stated in terms of $d\log(y_{t+1}/y_t)/d\log y_t$. We can write

$$\frac{d\log\frac{y_{t+1}}{y_t}}{d\log y_t} = \frac{\frac{d\left(\frac{y_{t+1}}{y_t}\right)}{\frac{y_{t+1}}{y_t}}}{\frac{dy_t}{y_t}} = \frac{y_t^2}{y_{t+1}} \frac{d\left(\frac{y_{t+1}}{y_t}\right)}{dy_t} = \frac{y_t^2}{y_{t+1}} \left(\frac{dy_{t+1}}{dy_t}\frac{1}{y_t} - \frac{y_{t+1}}{y_t^2} \right) = \frac{dy_{t+1}}{dy_t}\frac{1}{\gamma} - 1 = \frac{d\hat{y}_{t+1}}{d\hat{y}_t} - 1.$$

Thus, the sought regression coefficient is $\alpha + (1-\alpha)\frac{1-\delta}{\gamma} - 1$. Since $\alpha \in (0,1)$, this object lies in $(\frac{1-\delta-\gamma}{\gamma}, 0)$. Taking a period to be a year, one would set $\gamma = 0.02$ and $\delta = 0.10$, so with an $\alpha = 0.3$, roughly as in U.S. data, we obtain a coefficient of close to -0.08. Available regression estimates indicate a number less than half of this amount. I.e., the data suggests that a calibrated version of Solow's model implies convergence that is too fast.

Turning to a setting where s is chosen endogenously, the derivations above need to be complemented with an analysis of how a change in $\hat{k}_t$ (or $\hat{y}_t$) changes s. The analysis in Section 9.2.3 shows that, in the case without exogenous growth, $dy_{t+1}/dy_t = dk_{t+1}/dk_t$ at steady state is given by the smallest solution to

$$\lambda^2 - \left[1 + \frac{1}{\beta} + \frac{u'}{u''}\frac{f''}{f'}\right]\lambda + \frac{1}{\beta} = 0,$$

where the derivatives of u and f are evaluated at the steady-state point. The case with growth is straightforward to analyze, because in that case preferences do not influence the steady-state level of capital. So consider the case $u(c) = (1-\sigma)^{-1}(c^{1-\sigma} - 1)$. Study of the second-order polynomial equation yields that the lowest root decreases in the expression $\frac{u'}{u''}\frac{f''}{f'}$, which under the functional forms assumed (recall that $f(k) = k^\alpha + (1-\delta)k$) can be shown to become $\frac{1-\alpha}{\alpha\beta\sigma}(1-\beta(1-\delta))(1-\beta(1-\alpha\delta))$. Thus, a higher σ raises λ, making y move more slowly toward steady state. Intuitively, if there is significant curvature in utility, consumers do not like sharp movements in consumption, and since convergence precisely requires consumption to change, convergence will be slow. Slower convergence also follows from a high α, a high β, or a low δ.

9.3.2 Choosing to grow

The next issue to address is whether an individual who inhabits an economy in which there is some sort of exogenous technological progress, and in which the production technology is such that sustained growth is feasible, will choose a growing output path or not.

Initially, Solow overlooked this issue by assuming that capital accumulation rule was determined by the policy rule

$$i_t = sy_t,$$

where the savings rate $s \in [0,\, 1]$ was constant and exogenous. It is clear that such a rule can be consistent with a balanced growth path. Then the underlying premise is that the consumers' preferences are such that they choose a growing path for output.

However, this is too relevant an issue to be overlooked. What is the generality of this result? Specifically, what are the conditions on preferences for constant growth to obtain? Clearly, the answer is that not all types of preferences will work. We will restrict our attention to the usual time-separable preference relations. Hence the problem faced by a central planner will be of the form:

$$\max_{\{i_t,\, c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t u\left(c_t,\, n_t\right) \right\} \tag{9.5}$$

$$\begin{aligned} \text{s.t.} \quad & c_t + i_t = F\left(K_t,\, \gamma^t n_t\right) \\ & K_{t+1} = i_t + (1-\delta)\, K_t \\ & K_0 \text{ given.} \end{aligned}$$

For this type of preference relations, we have the following result:

Theorem 9.5 *Balanced growth is possible as a solution to the central planner's problem (9.6) if and only if*

$$u\left(c,\, n\right) = \frac{\left(cv(1-n)\right)^{1-\sigma} - 1}{1-\sigma},$$

where time endowment is normalized to one as usual and $v(\cdot)$ is a function with leisure as an argument.

Proving the theorem is rather endeavored in one of the two directions of the double implication, because the proof involves partial differential equations. Also notice that we say that balanced growth is a *possible* solution. The reason is that initial conditions also have an impact on the resulting output growth. The initial state has to be such that the resulting model dynamics (that may initially involve non-constant growth) eventually lead the system to a balanced growth path (constant growth). Arbitrary initial conditions do not necessarily satisfy this.

Comments:

1. Balanced growth involves a constant n.

2. $v(1-n) =$ constant fits the theorem assumptions; hence, non-valued leisure is consistent with balanced growth path.

3. What happens if we introduce a "slight" modifications to $u(c, n)$, and use a functional form like
$$u(c, n) = \frac{(c-\overline{c})^{1-\sigma} - 1}{1-\sigma} \quad ?$$
$\overline{c}$ can be interpreted as a minimum subsistence consumption level. When c gets large with respect to $\overline{c}$, risk aversion decreases. Then for a low level of consumption c, this utility function representation of preferences will not be consistent with a balanced growth path; but, as c increases, the dynamics will tend towards balanced growth. This could be an explanation to observed growth behavior in the early stages of development of poor countries.

9.3.3 Transforming the model

Let us now describe the steps of solving a model for a balanced growth path.

1) Assume that preferences are represented by the utility function
$$\frac{c^{1-\sigma}v(1-n)^{1-\sigma} - 1}{1-\sigma}.$$

2) Take first order conditions of the central planner's problem (9.6) described above using this preference representation.

3) Next *assume* that there is balanced growth, and show that the implied system of equations can be satisfied.

4) After solving for the growth rates transform the model into a stationary one.

We will perform these steps for the case of labor-augmenting technology under constant returns to scale. The original problem is
$$\max_{\{i_t, c_t, K_{t+1}, n_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma} v(1-n_t)^{1-\sigma} - 1}{1-\sigma} \right\} \tag{9.6}$$

$$\text{s.t.} \quad c_t + i_t = \gamma^t n_t F\left(\frac{K_t}{\gamma^t n_t}, 1\right)$$
$$K_{t+1} = i_t + (1-\delta) K_t$$
$$K_0 \text{ given.}$$

We know that the balanced growth solution to this Growth Model (9.7) has all variables growing at rate γ, except for labor. We define transformed variables by dividing each original variable by its growth rate:

$$\begin{aligned} \widehat{c}_t &= \frac{c_t}{\gamma^t} \\ \widehat{i}_t &= \frac{i_t}{\gamma^t} \\ \widehat{K}_t &= \frac{K_t}{\gamma^t}, \end{aligned}$$

and thus obtain the transformed model:

$$\max_{\left\{\widehat{i}_t, \widehat{c}_t, \widehat{K}_{t+1}, n_t\right\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t \frac{\widehat{c}_t^{1-\sigma} \gamma^{t(1-\sigma)} v\left(1-n_t\right)^{1-\sigma} - 1}{1-\sigma} \right\}$$

$$\text{s.t.} \quad \left(\widehat{c}_t + \widehat{i}_t\right) \gamma^t = \gamma^t n_t F\left(\frac{\widehat{K}_t \gamma^t}{\gamma^t n_t}, 1\right)$$
$$\widehat{K}_{t+1} \gamma^{t+1} = \left[\widehat{i}_t + (1-\delta) \widehat{K}_t\right] \gamma^t$$
$$K_0 \text{ given.}$$

Notice that we can write

$$\sum_{t=0}^{\infty} \beta^t \frac{\widehat{c}_t^{1-\sigma} \gamma^{t(1-\sigma)} v\left(1-n_t\right)^{1-\sigma} - 1}{1-\sigma} = \sum_{t=0}^{\infty} \widehat{\beta}^t \frac{\widehat{c}_t^{1-\sigma} v\left(1-n_t\right)^{1-\sigma} - 1}{1-\sigma} + \sum_{t=0}^{\infty} \widehat{\beta}^t \frac{1-\gamma^{-t(1-\sigma)}}{1-\sigma},$$

where $\widehat{\beta} = \beta \gamma^{(1-\sigma)}$. Then we can cancel out γ's to get:

$$\max_{\left\{\widehat{i}_t, \widehat{c}_t, \widehat{K}_{t+1}, n_t\right\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \widehat{\beta}^t \frac{\widehat{c}_t^{1-\sigma} v\left(1-n_t\right)^{1-\sigma} - 1}{1-\sigma} + \sum_{t=0}^{\infty} \widehat{\beta}^t \frac{1-\gamma^{-t(1-\sigma)}}{1-\sigma} \right\} \tag{9.7}$$

$$\text{s.t.} \quad \widehat{c}_t + \widehat{i}_t = n_t F\left(\frac{\widehat{K}_t}{n_t}, 1\right)$$
$$\widehat{K}_{t+1} \gamma = \widehat{i}_t + (1-\delta) \widehat{K}_t$$
$$K_0 \text{ given.}$$

Now we are back to the standard neoclassical growth model that we have been dealing with before. The only differences are that there is a γ factor in the capital accumulation equation, and the discount factor is modified.

We need to check the conditions for this problem to be well defined. This requires that $\beta\gamma^{1-\sigma} < 1$. Recall that $\gamma > 1$, and the usual assumption is $0 < \beta < 1$. Then:

1. If $\sigma > 1$, $\gamma^{1-\sigma} < 1$ so $\beta\gamma^{1-\sigma} < 1$ holds.

2. If $\sigma = 1$ then $\beta\gamma^{1-\sigma} = \beta < 1$ holds.

3. If $0 < \sigma < 1$, then for some parameter values of γ and β, we may run into an ill-defined problem.

Next we address the issue of the system behavior. If leisure is not valued and the production technology

$$f(k) \equiv F\left(\frac{K}{L}, 1\right) + (1-\delta)\frac{K}{L}$$

satisfies the Inada conditions ($f(0) = 0$, $f'(\cdot) > 0$, $f''(\cdot) < 0$, $\lim_{k\to\infty} f'(\cdot) = 0$, $\lim_{k\to 0} f'(\cdot) = \infty$) then global convergence to steady state obtains for the transformed model (9.7):

$$\lim_{t\to\infty} \widehat{c}_t = \overline{\widehat{c}},\ \lim_{t\to\infty} \widehat{i}_t = \overline{\widehat{i}},\ \lim_{t\to\infty} \widehat{k}_t = \overline{\widehat{k}}.$$

This is equivalent to saying that the original variables c_t, i_t, and k_t grow at rate γ asymptotically.

Therefore with the stated assumptions on preferences and on technology, the model converges to a balanced growth path, in which all variables grow at rate γ. This rate is exogenously determined; it is a parameter in the model. That is the reason why it is called "exogenous" growth model.

9.3.4 Adjustment costs and multisector growth models

- Convergence is influenced by adjustment costs.

- Consumption and investment sectors: special case with oscillations. Balanced growth.

- More sectors more generally: no general results.

- Structural change: agriculture, services, and manufacturing; can there be balanced growth?

- Other forms of changing technology.

9.4 Growth theory III: endogenous growth

The exogenous growth framework analyzed before has a serious shortfall: growth is not truly a result in such model - it is an assumption. However, we have reasons (data) to suspect that growth must be a rather more complex phenomenon than this long term productivity shift γ, that we have treated as somehow intrinsic to economic activity. In particular, rates of output growth have been very different across countries for long periods; trying to explain this fact as merely the result of different γ's is not a very insightful approach. We would prefer our model to produce γ as a result. Therefore, we look for endogenous growth models.

But what if the countries that show smaller growth rates are still in transition, and transition is slow? Could this be a plausible explanation of the persistent difference in growth? At least locally, the rate of convergence can be found from

$$\log y' - \log \overline{y} = \lambda \left(\log y - \log \overline{y} \right),$$

where λ is the eigenvalue smaller than one in absolute value found when linearizing the dynamics of the growth model (around the steady state). Recall it was the root to a second degree polynomial. The closer λ is to 1 (in absolute value), the slower the convergence. Notice that this equation can be rewritten to yield the growth regression:

$$\log y' - \log y = - \left(1 - \lambda\right) \log y + \left(1 - \lambda\right) \log \overline{y} + \alpha,$$

where $-(1-\lambda)$ is the β parameter in the growth regressions, $\log y$ shows up as $\log y_0$; $(1-\lambda)$ is the γ, and $\log \overline{y}$ is the residual z; finally α (usually called γ_0) is the intercept that shows up whenever a technological change drift is added.

In calibrations with "reasonable" utility and production functions, λ tends to become small in absolute value - hence not large enough to explain the difference in growth rates of e.g. Korea and Chad. In general, the less curvature the return function shows, the faster the convergence. The extreme special cases are:

1. u linear $\Rightarrow \lambda = 0$ - immediate convergence.
2. f linear $\Rightarrow \lambda = 1$ - no convergence.

The more curvature in u, the less willing consumers are to see their consumption pattern vary over time - and growth is a (persistent) variation. On the other hand, the more curvature in f, the higher the marginal return on capital when the accumulated stock is small; hence the more willing consumers are to put up with variation in their consumption stream, since the reward is higher.

9.4.1 The AK model

Let us recall the usual assumptions on the production technology in the neoclassical growth model: F was constant returns to scale, and also the "per capita" production function f

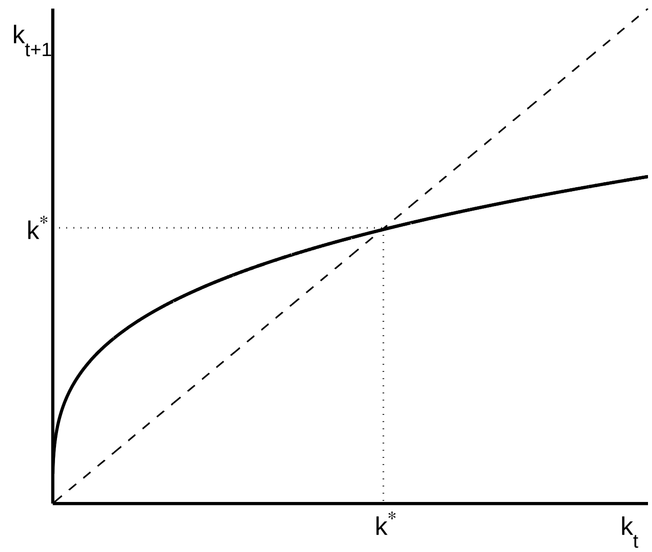

Figure 9.4: Global dynamics

satisfied: $f(0) = 0$, $f'(\cdot) > 0$, $f''(\cdot) < 0$, $\lim\limits_{x\to 0} f'(\cdot) = \infty$, and $\lim\limits_{x\to\infty} f'(\cdot) = 0$, with the global dynamics as depicted in Figure 9.4 (with a "regular" utility function).

Long run growth is not feasible. Notice that whenever the capital stock k exceeds the level k^*, then next period's capital will decrease: $k' < k$. In order to allow long run growth, we need the introduce at least some change to the production function: We must dispose of the assumption that $\lim\limits_{x\to\infty} f'(\cdot) = 0$. What we basically want is that f does not cross the 45^o line. Then $\lim\limits_{x\to\infty} f'(\cdot) > 0$ seems necessary for continuous growth to obtain.

If we have that $\lim\limits_{x\to\infty} f'(\cdot) = 1$ (that is, the production function is asymptotically parallel to the 45^o line), then exponential growth is not feasible - only arithmetic growth is. This means that we must have $\lim\limits_{x\to\infty} f'(\cdot) > 1$ for a growth *rate* to be sustainable over time.

The simplest way of achieving this is to assume the production technology to be represented by a function of the form:

$$f(k) = Ak$$

with $A > 1$. More generally, for any depreciation rate δ, we have that the return on capital is

$$\begin{aligned}
(1-\delta)\,k + f(k) &= (1-\delta)\,k + Ak \\
&= (1-\delta+A)\,k \\
&\equiv \widetilde{A}k,
\end{aligned}$$

so the requirement in fact is $A > \delta$ for exponential growth to be feasible (when $\delta < 1$).

The next question is whether the consumer will choose growth, and if so, how fast. We will answer this question assuming a CIES utility function (needed for balanced growth), with non-valued leisure. The planner's problem then is:

$$\begin{aligned} U = & \max_{\{c_t,\, k_{t+1}\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}}{1-\sigma} \right\} \\ \text{s.t.} \quad & c_t + k_{t+1} = A k_t, \end{aligned}$$

where $\sigma > 0$. The Euler Equation is

$$c_t^{-\sigma} = \beta c_{t+1}^{-\sigma} A.$$

Now we have that the growth rate of consumption must satisfy:

$$\frac{c_{t+1}}{c_t} = (\beta A)^{\frac{1}{\sigma}}.$$

The growth rate of consumption is a function of all the parameters in the utility function and the production function. Notice that this implies that the growth rate is constant as from $t = 0$. There are no transitional dynamics in this model; the economy is in the balanced growth path from the start. There will be long-run growth provided that

$$(\beta A)^{\frac{1}{\sigma}} > 1. \tag{9.8}$$

This does not quite settle the problem, though: an issue remains to be addressed. If the parameter values satisfy the condition for growth, is utility still bounded? We must evaluate the optimal path using the utility function:

$$U = \sum_{t=0}^{\infty} \left[\beta \left[(\beta A)^{\frac{1}{\sigma}} \right]^{1-\sigma} \right]^t \frac{c_0^{1-\sigma}}{1-\sigma}.$$

So the sufficient condition for boundedness is:

$$\beta \left[(\beta \alpha)^{\frac{1}{\sigma}} \right]^{1-\sigma} < 1. \tag{9.9}$$

The two conditions (9.8) and (9.9) must simultaneously hold for us to obtain a balanced growth path.

Remark 9.6 (Distortionary taxes and growth) *Notice that the competitive allocation in this problem equals the central planner's (why?). Now suppose that the government levies a distortionary tax on (per capita) capital income and uses the proceeds to finance a lump-sum transfer. Then the consumer's decentralized problem has the following budget constraint:*

$$c_t + k_{t+1} = (1 - \tau_k) R_t k_t + \tau_t,$$

while the government's budget constraint requires that

$$\tau_k R_t k_t = \tau_t.$$

This problem is a little more endeavored to solve due to the presence of the lump-sum transfers τ_t. Notwithstanding, you should know that τ_k (the distortionary tax on capital income) will affect the long run growth rate.

Remark 9.7 (Explanatory power) *Let us now consider how realistic the assumptions and the results of the model are:*

* **Assumptions** *The AK production function could be interpreted as a special case of the Cobb-Douglas function with $\alpha = 1$ - then labor is not productive. However, this contradicts actual data, that shows that labor is a hugely significant component of factor input. Clearly, in practice labor is important. But this is not captured by the assumed production technology.*

  We could imagine a model where labor becomes unproductive; e.g. assume that

  $$F_t\left(K_t,\, n_t\right) = AK_t^{\alpha_t} n_t^{1-\alpha_t}.$$

  Then if $\lim\limits_{t\to\infty} \alpha_t = 1$, we have asymptotic linearity in capital. But this is unrealistic.

* **Results** *The growth has become a function of underlying parameters in the economy, affecting preferences and production. Could the dispersion in cross-country growth rates be explained by differences in these parameters? Country i's Euler Equation (with a distortionary tax on capital income) would be:*

  $$\left(\frac{c_{t+1}}{c_t}\right)_i = \left[\beta_i A_i \left(1 - \tau_k^i\right)\right]^{\frac{1}{\sigma_i}}.$$

  But the problem with the AK model is that, if parameters are calibrated to mimic the data's dispersion in growth rates, the simulation results in too much divergence in output level. The dispersion in 1960-1990 growth rates would result in a difference in output levels wider than the actual.

Remark 9.8 (Transitional dynamics) *The AK model implies no transitional dynamics. However, we tend to see transitional dynamics in the data (recall the conditional convergence result in growth regressions).*

9.4.2 Romer's externality model

The intellectual precedent to this model is Arrow (1962). The basic idea is that there are externalities to capital accumulation, so that individual savers do not realize the full return on their investment. Each individual firm operates the following production function:

$$F\left(K,\, L,\, \overline{K}\right) = AK^{\alpha} L^{1-\alpha} \overline{K}^{\rho},$$

where K is the capital operated by the firm, and $\overline{K}$ is the aggregate capital stock in the economy. We assume that $\rho = 1 - \alpha$ so that in fact a central planner faces an AK-model decision problem. Notice that if we assumed that $\alpha + \rho > 1$, then balanced growth path would not be possible.

The competitive equilibrium will involve a wage rate equal to:

$$w_t = (1-\alpha)\, AK_t^{\alpha} L_t^{-\alpha} \overline{K}_t^{1-\alpha}.$$

Let us assume that leisure is not valued and normalize the labor endowment L_t to one in every t. Assume that there is a measure one of representative firms, so that the equilibrium wage must satisfy

$$w_t = (1-\alpha)\, A\overline{K}_t.$$

Notice that in this model, wage increases whenever there is growth, and the wage as a fraction of total output is substantial. The rental rate, meanwhile, is given by:

$$R_t = \alpha A.$$

The consumer's decentralized Euler Equation will be (assuming a CIES utility function and $\delta = 1$):

$$\frac{c_{t+1}}{c_t} = (\beta R_{t+1})^{\frac{1}{\sigma}}.$$

Substituting for the rental rate, we can see that the rate of change in consumption is given by:

$$g_c^{CE} = (\beta\alpha A)^{\frac{1}{\sigma}}.$$

It is clear that since a planner faces an AK model his chosen growth rate should be:

$$g_c^{CP} = (\beta A)^{\frac{1}{\sigma}}.$$

Then $g_c^{CP} > g_c^{CE}$: the competitive equilibrium implements a lower than optimal growth rate, which is consistent with the presence of externalities to capital accumulation.

Remark 9.9 (Pros and cons of this model) *The following advantages and disadvantages of this model can be highlighted:*

\+ *The model overcomes the "labor is irrelevant" shortfall of the AK model.*

− *There is little evidence in support of a significant externality to capital accumulation. Notice that if we agreed for example that $\alpha = 1/3$, then the externality effect would be immense.*

− *The model leads to a large divergence in output levels, just as the AK model.*

9.4.3 Human capital accumulation

Now let "labor hours" in the production function be replaced by "human capital". Human capital can be accumulated, so the technology does not run into decreasing marginal returns. For example, in the Cobb-Douglas case, we have:

$$F\left(K,\, H\right) = AK^{\alpha}H^{1-\alpha}.$$

There are two distinct capital accumulation equations:

$$\begin{aligned} H_{t+1} &= \left(1-\delta^{H}\right) H_t + I_t^H \\ K_{t+1} &= \left(1-\delta^{K}\right) K_t + I_t^K, \end{aligned}$$

and the resource constraint in the economy is:

$$c_t + I_t^H + I_t^K = AK_t^{\alpha}H_t^{1-\alpha}.$$

Notice that, in fact, there are two assets: H and K. But there is no uncertainty; hence one is redundant. The return on both assets must be equal.

Unlike the previous model, in the current setup a competitive equilibrium does implement the central planner's solution (why can we say so?). Assuming a CIES utility function and a general production function $F(\cdot,\cdot)$, the first order conditions in the central planner's problem are:

$$\begin{aligned} c_t &: \beta^t c_t^{-\sigma} = \lambda_t \\ K_{t+1} &: \lambda_t = \lambda_{t+1}\left[1-\delta^K + F_K\left(K_{t+1},\, H_{t+1}\right)\right] \\ H_{t+1} &: \lambda_t = \lambda_{t+1}\left[1-\delta^H + F_H\left(K_{t+1},\, H_{t+1}\right)\right], \end{aligned}$$

which leads us to two equivalent instances of the Euler Equation:

$$\frac{c_{t+1}}{c_t} = \left(\beta\left[1-\delta^K + F_K\left(\frac{K_{t+1}}{H_{t+1}},\, 1\right)\right]\right)^{\frac{1}{\sigma}} \tag{9.10}$$

$$\frac{c_{t+1}}{c_t} = \left(\beta\left[1-\delta^H + F_H\left(\frac{K_{t+1}}{H_{t+1}},\, 1\right)\right]\right)^{\frac{1}{\sigma}}. \tag{9.11}$$

Notice that if the ratio $\frac{K_{t+1}}{H_{t+1}}$ remains constant over time, this delivers balanced growth. Let us denote $x_t \equiv \frac{K_t}{H_t}$. Then we have

$$1-\delta^K + F_K\left(x_t,\ 1\right) = 1-\delta^H + F_H\left(x_t,\ 1\right). \tag{9.12}$$

But then the equilibrium in the asset market requires that $x_t = \overline{x}$ be constant for all t (assuming a single solution to (9.12)); and $\overline{x}$ will depend only on δ^H, δ^K, and parameters of the production function F.

Example 9.10 *Assume that* $\delta^H = \delta^K$*, and* $F(K, H) = AK^\alpha H^{1-\alpha}$*. Then since RHS of (9.10) must equal RHS of (9.11) we get:*

$$\alpha A x^{\alpha-1} = (1-\alpha) A x^\alpha$$

$$\Rightarrow x = \frac{\alpha}{1-\alpha} = \frac{K_t}{H_t}.$$

From $t = 1$ *onwards,* $K_t = xH_t$*. Then*

$$\begin{aligned} AK_t^\alpha H_t^{1-\alpha} &= A (xH_t)^\alpha H_t^{1-\alpha} \\ &= \widetilde{A} H_t \\ &= \widehat{A} K_t, \end{aligned}$$

where $\widetilde{A} \equiv Ax^\alpha$*, and* $\widehat{A} \equiv Ax^{1-\alpha}$*. In any case, this reduces to an* AK *model.*

Remark 9.11 (Pros and cons of this approach) *We can highlight the following advantages and disadvantages of this model:*

\+ *Labor is treated seriously, and not resorting to "tricks" like externalities.*

− *The law of motion of human capital is too mechanistic:*

$$H_{t+1} = \left(1 - \delta^H\right) H_t + I_t^H.$$

Arguably, knowledge might be bounded above at some point. This issue could be counter-argued by saying that H_t *should be interpreted as general formation (such as on-the-job training, etcetera), and not narrowly as schooling.*

LUCAS'S MODEL HERE

− *This model implies divergence of output levels; it is an* AK *model in essence.*

9.4.4 Endogenous technological change

Product variety expansion

Based on the Cobb-Douglas production function $F(K, L) = AK^\alpha L^{1-\alpha}$, this model seeks to make A endogenous. One possible way of modelling this would be simply to make firms choose the inputs knowing that this will affect A. However, if A is increasing in K and L, this would lead to increasing returns, since for any $\lambda > 1$

$$A(\lambda K, \lambda L)(\lambda K)^\alpha (\lambda L)^{1-\alpha} > \lambda A K^\alpha L^{1-\alpha}.$$

An alternative approach would have A being the result of an external effect of firm's decisions. But the problem with this approach is that we want A to be *somebody's* choice; hence, an externality will not work.

One way out of this dilemma is to drop the assumption of perfect competition in the economy. In the model to be presented, A will represent "variety" in production inputs. The larger A, the wider the range of available production (intermediate) goods. Specifically, let capital and consumption goods in this economy be produced according to the function

$$y_t = L_t^\beta \int_0^{A_t} x_t^{1-\beta}(i)\, di,$$

where i is the type of intermediate goods, and $x_t(i)$ is the amount of good i used in production at date t. Therefore, there is a measure A_t of different intermediate goods. You may notice that the production function exhibits constant returns to scale.

The intermediate goods $x_t(i)$ are produced with capital goods using a linear technology:

$$\int_0^{A_t} \eta x_t(i)\, di = K_t,$$

i.e., η units of capital are required to produce 1 unit of intermediate good of type i, for all i.

The law of motion and resource constraint in this economy are the usual:

$$\begin{aligned} &K_{t+1} = (1-\delta)\, K_t + I_t \\ &c_t + I_t = y_t. \end{aligned}$$

We will assume that an amount L_{1t} of labor is supplied to the final goods production sector at time t. In addition, we temporarily assume that A_t grows at rate γ (since growth in A_t is actually endogenous):

$$A_{t+1} = \gamma A_t.$$

Given this growth in A, is long run output growth feasible? The key issue to answer this question is to determine the allocation of capital among the different types of intermediate goods. Notice that this decision is of a static nature: the choice at t has no (dynamic) consequences on the future periods' state. So the production maximizing problem is to:

$$\begin{aligned} &\max_{x_t(i)} \left\{ L_{1t}^\beta \textstyle\int_0^{A_t} x_t^{1-\beta}(i)\, di \right\} \\ &\text{s.t.} \ \ \textstyle\int_0^{A_t} \eta x_t(i)\, di = K_t. \end{aligned}$$

Since the objective function is concave, the optimal choice has $x_t(i) = x_t$ for all i. This outcome can be interpreted as a preference for "variety" - as much variety as possible is chosen.

Substituting the optimal solution in the constraint:

$$\begin{aligned} \int_0^{A_t} \eta x_t\, di &= K_t \\ A_t x_t \eta &= K_t. \end{aligned} \tag{9.13}$$

Maximized production is:

$$\begin{aligned} y_t &= L^{\beta} \int_0^{A_t} x_t^{1-\beta}\, di \\ &= L^{\beta} A_t x_t^{1-\beta}. \end{aligned} \tag{9.14}$$

Using (9.13) in (9.14),

$$\begin{aligned} y_t &= L_{1t}^{\beta} A_t \left(\frac{K_t}{\eta A_t} \right)^{1-\beta} \\ &= \frac{L_{1t}^{\beta}}{\eta^{1-\beta}} A_t^{\beta} K_t^{1-\beta}. \end{aligned}$$

Clearly A_t^{β} grows if A_t grows at rate γ. If we conjecture that K_t also grows at rate γ, then the production function is linear in the growing terms. Therefore, the answer to our question is "yes": a balanced growth path *is* feasible; with K_t, y_t and A_t growing at rate γ.

The next issue is how to determine γ, since we are dealing with an endogenous growth model. We will make the following assumption on the motion equation for A_t:

$$A_{t+1} = A_t + L_{2t} \delta A_t,$$

where L_{2t} denotes labor effort in research and development, and $L_{2t}\delta$ is the number of new "blueprints" that are developed at time t, as a consequence of this R&D. This motion equation resembles a learning by doing effect.

Exercise 9.12 *Let the consumer have the standard CIES preferences*

$$U(c) = \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma} - 1}{1-\sigma}.$$

Assume that leisure is not valued, and total time endowment is normalized to 1. *Then the amount of labor effort allocated to the production and to the R&D sectors must satisfy the constraint:*

$$L_{1t} + L_{2t} = 1.$$

Solve the planning problem to obtain (an equation determining) the balanced growth rate γ.

The decentralized economy

We will work with the decentralized problem. We assume that there is perfect competition in the final output industry. Then a firm in that industry solves at time t:

$$\max_{x_t(i),\, L_{1t}} \left\{ L_{1t}^{\beta} \int_0^{A_t} x_t^{1-\beta}(i)\ di - w_t L_{1t} - \int_0^{A_t} q_t(i)\, x_t(i)\ di \right\}.$$

Notice that the firm's problem is a static one - w_t and $q_t(i)$ are taken as given. Equilibrium in the final goods market then requires that these are:

$$\begin{aligned} w_t &= \beta L_{1t}^{\beta-1} \int_0^{A_t} x_t^{1-\beta}(i)\, di \\ q_t(i) &= (1-\beta) L_{1t}^{\beta} x_t^{-\beta}(i)\,. \end{aligned} \tag{9.15}$$

As for the intermediate goods industry, instead of perfect, we will assume that there is *monopolistic competition*. There is only *one* firm per type i (a patent holder). Each patent holder takes the demand function for its product as given. Notice that (9.15) is just the inverse of this demand function. All other relevant prices are also taken as given - in particular, the rental rate R_t paid for the capital that is rented to consumers. Then the owner of patent i solves:

$$\begin{aligned} \pi(i) &= \max_{K_t^i} \left\{ q_t(i)\, x_t(i) - R_t K_t^i \right\} \\ & \quad \text{s.t. } x(i)\,\eta = K_t^i, \end{aligned} \tag{9.16}$$

or equivalently, using (9.15) and (9.16),

$$\pi(i) = \max_{K_t^i} \left\{ (1-\beta) L_{1t}^{\beta} \left(\frac{K_t^i}{\eta} \right)^{1-\beta} - R_t K_t^i \right\}.$$

The first-order conditions for this problem are:

$$(1-\beta)^2 L_{1t}^{\beta} \eta^{\beta-1} \left(K_t^i\right)^{-\beta} = R_t.$$

Observe that $\pi(i) > 0$ is admissible: the firm owns a patent, and obtains a rent from it. However, this patent is not cost free. It is produced by "R&D firms", who sell them to intermediate goods producers. Let p_t^P denote the price of a patent at time t. Then ideas producers solve:

$$\begin{aligned} &\max_{A_{t+1},\, L_{2t}} \left\{ p_t^P \left(A_{t+1} - A_t\right) - w_t L_{2t} \right\} \\ &\quad \text{s.t. } A_{t+1} = A_t + L_{2t} \delta A_t. \end{aligned}$$

We will assume that there is free entry in the ideas industry. Hence, there must be zero profits from engaging in research and development. Notice that there is an externality (sometimes called "standing on the shoulders of giants"). The reason is that the decision involving the change in A, $A_{t+1} - A_t$, affects production at $t+j$ via the term δA_{t+j} in the equation of motion for A_{t+j}. But this effect is not realized by the firm who chooses the change in A. This is the second reason why the planner's and the decentralized problems will have different solutions (the first one was the monopoly power of patent holders).

The zero profit condition in the ideas industry requires that the price p_t^P be determined from the first-order condition

$$p_t^P \delta A_t = w_t,$$

where w_t is the same as in the market for final goods.

Once this is solved, if p_t^C denotes the date-0 price of consumption (final) goods at t, then we must have

$$p_t^P p_t^C = \sum_{s=t+1}^{\infty} \pi_s (i) \, p_s^C .$$

As a result, *nobody makes profits in equilibrium.* The inventors of patents appropriate the extraordinary rents that intermediate goods producers are going to obtain from purchasing the rights on the invention.

Balanced growth

Next we solve for a (symmetric) balanced growth path. We assume that all variables grow at (the same, and) constant rates:

$$\begin{aligned} K_{t+1} &= \gamma K_t \\ A_{t+1} &= \gamma A_t \\ c_{t+1} &= \gamma c_t \\ L_{1t} &= L_1 \\ L_{2t} &= L_2 \\ w_{t+1} &= \gamma w_t . \end{aligned}$$

With respect to the intermediate goods $x_t (i)$, we already know that an equal amount of each type of them is produced each period: $x_t (i) = x_t$. In addition, we have that this amount must satisfy:

$$A_t \eta x_t = K_t .$$

Since both A_t and K_t (are assumed to) grow at rate γ, then x_t must remain constant for this equation to hold for every t. Hence,

$$x_t = x = \frac{K_t}{A_t \eta} .$$

Then the remaining variables in the model must remain constant as well:

$$\begin{aligned} R_t &= R \\ \pi_t (i) &= \pi \\ p_t^P &= p^P \\ q_t (i) &= q . \end{aligned}$$

It is up to you to solve this problem:

Exercise 9.13 *Given the assumptions on consumer's preferences as in exercise 9.12, write down a system of n equations and n unknowns determining γ, L_1, L_2, etc. After that, compare the growth rate in decentralized economy with the planner's growth rate γ which you have already found. Which one is higher?*

Product ladders

AGHION-HOWITT STYLE SETTINGS, BUT WITHOUT SOLVING. INVESTMENT-SPECIFIC MODEL.

9.4.5 Directed technological change

9.4.6 Models without scale effects

9.5 What explains long-run growth and the world income distribution?

9.5.1 Long-run U.S. growth

9.5.2 Assessing different models

Based on the observation that we have not seen a large fanning out of the distribution of income over countries it is hard to argue that an endogenous-growth model, with countries growing at different rates in the long run, is approximately right. Thus, the key is to find a model of (i) relative income levels and (ii) world growth. In the discussion below, we will focus on the former, though making some brief comments on the latter.

The degree of convergence

One of the key elements of testing the explanatory power of both the exogenous and the endogenous growth models is their implications regarding convergence of growth rates across different countries. Recall the sixth of the Kaldor's stylized facts: growth rates are *persistently* different across countries. The models discussed above imply:

Exogenous growth	*vs.*	Endogenous growth
$AK^{\alpha}L^{1-\alpha}$		AK
does not lead to divergence.		leads to divergence in relative income levels.

Is it possible to produce divergence (or at least slow convergence) in the exogenous growth framework through appropriate calibration? Using $\alpha = 1/3$, the exogenous growth model leads to too fast convergence. A "brilliant" solution is to set $\alpha = 2/3$. The closer to 1 α is set, the closer is the exogenous growth model to the AK model.

However, we are not so free to play around with α. This parameter can be measured from the data:

$$\alpha = \frac{KF_K}{y} = \frac{KR}{y}.$$

A possible solution to this problem is to introduce a "mystery capital", S, so that the production function looks like:

$$y = AK^{\alpha}L^{\beta}S^{1-\alpha-\beta}.$$

Or, alternatively introduce "human capital" as the third production factor, besides physical capital and labor:

$$y = AK^{\alpha}L^{\beta}H^{1-\alpha-\beta}.$$

Income differences and investment return differences

We will explore the argument developed by Lucas to study the implications of the growth model for cross-country differences in rates of return on investment. This will allow us to study how actual data can be used to test implications of theoretical models.

There is a significant assumption made by Lucas: suppose that it was possible to export U.S. production technology (or "know how") to other countries. Then the production function, both domestically and abroad, would be

$$y = AK^{\alpha}L^{1-\alpha}$$

with a different level of K and L in each country, but the same A, α, and capital depreciation level δ. Then imagine a less developed country whose annual (per capita) output is a seventh of the US output:

$$\frac{y_{LDC}}{y_{US}} = \frac{1}{7}. \tag{9.17}$$

Using per capita variables ($L_{US} = L_{LDC} = 1$), the marginal return on capital investment in the US is calculated as:

$$R_{US} = \alpha AK_{US}^{\alpha-1} - \delta,$$

where the parameters α and δ take values of 1/3 and .1, respectively.

The net rate of return on capital in the US can be estimated to be 6.5% per annum, so the net rate is:

$$R_{US} = 0.065.$$

Manipulating the Cobb-Douglas expression a little,

$$\alpha AK_{US}^{\alpha-1} = \alpha\frac{AK_{US}^{\alpha}}{K_{US}} = \alpha\frac{y_{US}}{K_{US}}.$$

What is the return on capital in the less developed country?

$$R_{LDC} = \alpha\frac{y_{LDC}}{K_{LDC}} - \delta.$$

We have that

$$7 = \frac{y_{US}}{y_{LDC}} = \frac{AK_{US}^{\alpha}}{AK_{LDC}^{\alpha}} = \left(\frac{K_{US}}{K_{LDC}}\right)^{\alpha}. \tag{9.18}$$

So, from (9.17) and (9.18),

$$\frac{y_{LDC}}{K_{LDC}} = \frac{7^{-1} \cdot y_{US}}{7^{-\frac{1}{\alpha}} \cdot K_{US}} = 7^{\frac{1-\alpha}{\alpha}} \cdot \frac{y_{US}}{K_{US}},$$

and, using $\alpha = 1/3$,

$$\frac{y_{LDC}}{K_{LDC}} = 7^2 \cdot \frac{y_{US}}{K_{US}}.$$

We know from the data that

$$.065 = \frac{1}{3} \cdot \frac{y_{US}}{K_{US}} - .1$$
$$\Rightarrow \frac{y_{US}}{K_{US}} = .495.$$

Therefore,

$$\frac{y_{LDC}}{K_{LDC}} = 49 \cdot \frac{y_{US}}{K_{US}} = 49 \cdot .495 = 24.255,$$

which implies that the (net) rate of return on capital in the less developed country should be:

$$R_{LDC} = \frac{1}{3} \cdot 24.255 - .1 = 7.985.$$

This is saying that *if the US production techniques could be exactly replicated* in less developed countries, the net return on investment would be 798.5%. This result is striking since if this is the case, then capital should be massively moving out of the US and into less developed countries. Of course, this riddle might disappear if we let $A_{LDC} < A_{US}$[2].

Exercise 9.14 *Assume that rates of return on capital are in fact the same in the US and in LDC. Assume also that* $\alpha_{US} = \alpha_{LDC}$*,* $\delta_{US} = \delta_{LDC}$*, but* $A_{US} \neq A_{LDC}$*. Under these assumptions, what would be the difference in capital and total factor productivity (A) between the US and LDC given that* $\frac{y_{US}}{y_{LDC}} = 7$*?*

HUMAN CAPITAL EXTERNALITIES HERE

Other ideas

9.5.3 Productivity accounting

9.5.4 A stylized model of development

References

[2]The calculation also assumes away the differences in riskiness of investment, transaction costs, etc.

Arrow K., “The economic implications of learning by doing,” *Review of Economic Studies*, 29(1962), 155-173.

Greenwood, Jeremy, Zvi Hercowitz and Per Krusell, “Long-Run Effects of Investment-Specific Technological Change,” *American Economic Review*, 87(1997), 342-362.

Lucas, Robert E. Jr., “Why doesn't capital flow from rich countries to poor countries?” *American Economic Review*, 80(1990), (Volume 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association), 92-96.

Chapter 10

Business cycles

The purpose of this chapter is to introduce the study of business cycles. By business cycles we mean fluctuations of output around its long term growth trend. In this sense, this chapter complements growth theory to provide a thorough explanation of the behavior of economic aggregates: first, output grows secularly; second, it fluctuates around its long term trend. We have already analyzed the former phenomenon. The latter is our topic now.

We will first overview the empirical facts that constitute our object of study, and the history of the attempts at explaining these facts. After that, we will study the theory that has developed. We could separate the evolution of this theory in three phases: (i) Pre-Real Business Cycles; (ii) Real Business Cycles; (iii) Current trends.

10.1 Introduction

There are two basic questions that gave birth to this area of macroeconomics:

1. Why are there cycles?
2. How do they work?

Before the real business cycle revolution, Keynesianism approached the understanding of the business cycle by postulating that investors were driven by "animal spirits". These non-rational feelings of investors propelled them to frantically invest or gloomily refrain from doing so, according to the prevailing mood. This notion has not completely disappeared from economics, however elaborate the explanations of investor behavior have now come to be. The current version of "animal spirits" does not refer to the moods of those who make investment decisions, but of those who make consumption decision: it is the fashionable indicator of consumer confidence. Apparently, people go to mall whenever they wake up feeling confident.

The Keynesians and their intellectual heirs did not base their approach to the business cycles on micro-foundations of macroeconomic behavior. Quite on the contrary, they study the effects of the above-mentioned moods on aggregate variables such as output and employment. Since acting on moods is an irrational way to make decisions, the economy looses potential value due to this lack of rationality; hence the government is called upon to correct this behavior. Therefore, the role of the government is one of the main topics of interest for these traditions in macroeconomics.

However, Lucas' critique (Lucas (1976)) of the aggregative approach to macroeconomics and more importantly Lucas (1977) generated the real business cycle revolution. The pioneering works in this tradition were Kydland and Prescott (1982), and Long and Plosser (1983). Besides its relevance and ability to explain the business cycle, this approach has had a very significant methodological impact on the practice of macroeconomics.

According to this view, the reason for cycles in the economy is that there are technology shocks that affect the productivity of factors. The source of the shock is real, and the propagation mechanism is real as well: it is a consequence of the intertemporal substitution of labor that optimizing decision makers choose whenever confronted with such a technology shock.

The critique to this approach is that the definition of a technology shock is somewhat blurred. What is exactly such a shock? Notwithstanding this weakness, the real business cycle tradition has data against which to contrast its hypotheses. Technology shocks can be measured through the de-trended Solow residual from actual data.

Finally, the reason why this tradition has focused on the "real" explanation of business cycles is rather accidental. When Prescott undertook the research program laid down by Lucas (1977) paper, the initial schedule was to start with a real source of the cycle (the technology shock) and the real propagation mechanism (the inter-temporal substitution), thereafter to increase the complexity of the model and allow for monetary mechanisms. However, on seeing that the real approach was providing a seemingly thorough explanation of the phenomenon, the course of the research program deviated towards increasing the complexity and richness of the real setup (such as introducing heterogeneity in agents, incomplete markets, etc.).

Of course, the real business cycle tradition is not the only one claiming the ability to provide an explanation of short run fluctuations of output around its growth trend. Among the main current contestants, the most relevant are:

(i) *New Keynesian models*. Opposed to the real approach, these take a monetary approach: The source of the cycles are monetary fluctuations, and the main propagation mechanism is also monetary: price "stickiness".

(ii) *Sunspot theories*. These are micro foundations models in which agents have full rationality, but are faced with models that have multiple equilibria. This allows for self-fulfilling, rational expectations that may cause fluctuations of output, even in spite of the absence of an underlying change in the production or utility fundamentals in the economy. This can be interpreted as a coordination problem, in which agents fail to achieve the "best" equilibrium out of the available ones. Notice that to some extent, the "animal spirits" (or consumer confidence) concept can be accommodated to explain why agents simultaneously believe that a given equilibrium will be realized, and act accordingly.

Before embarking on our topic of study, let us make a final comment on the current state of the art, in particular of the real approach. Most of the research has modeled typical, complete markets, usually operating under perfect competition. This rules out the possibility of the structure of markets itself playing a role in the business cycle. Notice that in the case of the New Keynesians, this is quite the opposite: it is the structure of a single market (the money market) which generates and propagates the fluctuations. Without taking this rather extreme view, the real approach could be enriched by allowing the structure of markets to have its share of the cycle phenomenon. The new literature is exploring this by introducing in the behavior of decision makers the need to "search". Information on prices and employment opportunities are not immediately available, as in the typical decentralized version of the planner's problem as we have studied it. Introducing the possibility of informational frictions in markets can account for the existence of unemployment, and give a role to money in the business cycle.

10.2 Stylized facts

In this section we are interested in presenting the main "facts" that business cycle theory seeks to explain. We take a rather epistemological definition of the word: By "Facts" we mean not exactly data, but rather what the economics profession regards as the acceptable indicators to be drawn from that data, and what the meaning is. The history of business cycles research has a very significant "dialectic" dimension: What are the "facts" to be explained? How should these be presented? What is the correct methodology to transform raw data into acceptable "facts"? All these questions are more than just methodological: they also reveal the different underlying visions of the economic phenomenon that are sustained by different schools in the profession. In the (extremely) brief overview of the history of "facts" that follows, this underlying debate can be seen to take the form of a methodological discussion.

10.2.1 Early Facts

The first intellectual precedent in the study of short run fluctuations in output was Burns and Mitchell (1946). Their purpose was to obtain stylized facts, à la Kaldor's growth facts. Their main findings were:

- Output in different sectors of the economy have positive covariance.

- Both investment and consumption of durable goods exhibit high variability.

- Company profits are very pro-cyclical and variable.

– Prices are pro-cyclical as well. (This is not true for the post-war years, but it was for the sample analyzed by Burns and Mitchell.)

– The amount of outstanding money balances and the velocity of circulation are pro-cyclical.

– Short term interest rates are pro-cyclical.

– Long term interest rates are counter-cyclical.

– Business cycles are "all alike" across countries and across time.

Burns and Mitchell's work was harshly criticized by Koopmans (1947). This critique was rather of a statistical, methodological nature. The main weaknesses highlighted in Burns and Mitchell's research were that:

- The work was not carefully done, and was hard to replicate.

- There was no solid underlying statistical theory. Relevant issues were not addressed altogether, such as the statistical significance of the assertions.

Koopmans' counter-argument discredited Burns and Mitchell's approach to the extent that no literature developed to improve and further their view. Instead of this, the leading study of business cycles was undertaken in Yale's Cowles commission, which consisted of studying huge econometric models of macroeconomic variations. This was called the "macroeconometrics" approach. The main authors in this tradition were Klein (Nobel prize due to this research) and Goldberg.

However, little has been left behind by this methodology, which ended up consisting of building up large scale macroeconomic models, making them bigger and bigger variable-wise until the regressions explained something. Finally, Lucas' critique (Lucas (1976)), that found widespread agreement through the economic profession, put an end to this research program.

As a result, research found itself needy of new stylized facts to explain since regressions were no longer regarded as a valid phenomenological source. The task to provide for credible (maybe just properly presented!) stylized facts, and then a suitable theoretical framework to explain them, was undertaken by Kydland and Prescott.

10.2.2 Kydland and Prescott (1982): How to convert raw data into facts

Kydland and Prescott's work is the founding stone of the current consensus on what "facts" are. These authors went back to the Burns and Mitchell tradition of stylized facts. Unlike their predecessors they succeeded because they were able to provide a solid methodological foundation for their approach.

In the first place, since the phenomenon to be studied is the short-run fluctuations of output around its long-term growth, these fluctuations need to be pinned down with precision. Raw data need to be rid of the secular growth component before the cycle can be identified. This is done by filtering the data, using the method developed by Hodrick and Prescott (the so-called "HP filter").

The HP filter works in the following way. Given a time series y_t, the purpose is to find out the trend component $\overline{y}_t$, and with this to calculate the value of the residual $y_t - \overline{y}_t$. This residual will be the data from which "facts" will be drawn.

The HP filter procedure to de-trend data is to solve the following minimization problem:

$$\min_{\{\overline{y}_t\}_{t=1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - \overline{y}_t)^2 \right\}$$
$$s.t. \sum_{t=2}^{T-1} \left[(\overline{y}_{t+1} - \overline{y}_t) - (\overline{y}_t - \overline{y}_{t-1}) \right]^2 \leq K.$$

In practice, K is set equal to 0, and this leads to the following Lagrangian:

$$L = \sum_{t=2}^{T-1} \left\{ (y_t - \overline{y}_t)^2 - \mu \left[(\overline{y}_{t+1} - \overline{y}_t) - (\overline{y}_t - \overline{y}_{t-1}) \right]^2 \right\} + (y_T - \bar{y}_T)^2 + (y_1 - \bar{y}_1)^2.$$

Hodrick and Prescott chose $\mu = 1600$ to de-trend quarterly data, and $\mu = 400$ for annual data. Once the problem is solved, the object of study is the resulting $\{y_t - \overline{y}_t\}_{t=1}^{T}$ sequence. With this in hand, "facts" in business cycles research are a series of relevant statistics computed from de-trended data.

10.2.3 Kydland and Prescott's facts

1. Volatilities

   Given a variable x, we define its percentage standard deviation as:

   $$\sigma_x \equiv \frac{(Var(x))^{1/2}}{\mu_x},$$

   where μ_x denotes the mean value of x.

   Then we have the following findings:

   - $\sigma_c < \sigma_y$,

     where $c \equiv$ consumption and $y \equiv$ output.

     What's behind this fact? Why is consumption less volatile than output? This can be interpreted as evidence for consumption smoothing behavior by agents.

   - $\sigma_{c_o} > \sigma_y$,

     where $c_o \equiv$ consumer durables.

   - $\sigma_i \approx 3 \cdot \sigma_y$,

     where $i \equiv$ investment.

- $\sigma_{TB} > \sigma_y$,

  where $TB \equiv$ trade balance.
- $\sigma_N \approx \sigma_y$,

  where $N \equiv$ total hours worked.
- $\sigma_E \approx \sigma_y$,

  where $E \equiv$ employment.
- $\sigma_{N/week} < \sigma_y$,

  where $N/week \equiv$ hours per week.
- $\sigma_K \ll \sigma_y$,

  where $K \equiv$ capital stock.

  In short-term periods, the stock of capital exhibits little variation.
- $\sigma_w < \sigma_{y/N}$,

  where $w \equiv$ real wage $=$ marginal product of labor and $y/N \equiv$ output per worked hour, i.e. labor productivity.

  The implication of this finding is that real wages are "sticky" - there is some smoothing of real wage fluctuations.

2. Correlations

- $\rho\left(\frac{y}{N}, y\right) > 0.$
- $\rho\left(w, y\right) \approx 0.$

  Recall that $y\,/\,N$ is the average product of labor, and w is the marginal product.
- $\rho\left(K, y\right) \approx 0.$
- $\rho\left(P, y\right) < 0$ (in post-war period),

  where $P \equiv$ price level.

3. Persistence

- $\rho\left[\left(y_t - \overline{y}_t\right), \left(y_{t-1} - \overline{y}_{t-1}\right)\right] \approx 0.9$ (from quarterly data).

4. Leads and Lags

   This addresses questions such as whether consumption leads output or investment leads output. No strong patterns were found on this regard by Kydland and Prescott.

5. The Business Cycle

   Finally, to top off the paradigm debate that underlies the methodological discussion, we must mention that the word "cycles" was not used in Kydland and Prescott (1982). What the authors sought to study were volatilities and correlations in economic variables, not "cycles". Nevertheless, the NBER has a cycle-measuring methodology that assigns beginning and ending dates to business cycles.

10.3 Real business cycle theory: the basic model

10.3.1 Basic methodology: Introduction to calibration

Once a definition of "facts" is at hand, a theory to account for them can be developed. The research on this was initiated by Kydland and Prescott (1982) and Long and Plosser (1983). The framework is the stochastic neoclassical growth model. And, remember: this project is quantitative. Everything is in numbers. The success of a real business cycle model is measured by its ability to numerically replicate the "facts".

The basic model is the central planner's problem to optimize the use of resources according to a time-additive preference relation that admits a utility function representation. For example, if production is affected by a shock on total factor productivity that follows an AR(1) process, the problem is:

$$\max_{\{c_t,\, n_t,\, l_t,\, K_{t+1}\}_{t=0}^{\infty}} \left\{ E_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t,\, l_t\right) \right] \right\}$$

$$\begin{aligned} \text{s.t.}\quad & c_t + x_t = z_t F\left(K_t,\, n_t\right) \\ & K_{t+1} = \left(1-\delta\right) K_t + x_t \\ & l_t + n_t = 1 \\ & z_{t+1} = \rho z_t + \varepsilon_{t+1}. \end{aligned}$$

The central methodological issue is how to pick the parameters in the utility and production functions. In this sense, the work of Kydland and Prescott has also a dialectic dimension. The authors are advocates of the technique known as "calibration". This is more than merely how to pick values for parameters to solve a numerical problem. It is a way of contrasting models against data as opposed to traditional econometrics.

Calibration, sometimes also called "back-of-the-envelope calculations", requires that values for parameters be picked from sources independent of the phenomenon under study. The discipline advocated by Kydland and Prescott bans "curve fitting" practices. For example, admissible sources of parameter values are:

- Household data on consumption, hours worked, and other microeconomic evidence, for individual preference parameters.

- Long run trend data for the factor shares in production (namely α in the Cobb-Douglas case).

10.3.2 Measurement: Solow growth accounting

The hypothesis of Kydland and Prescott (1982) and Long and Plosser (1983) is that the source of the observed volatilities and correlations in de-trended variables is a "technology shock". This is an aggregate stochastic shock that affects production, for example through total factor productivity as laid down above. There might be other types of stochastic shocks, however, such as changes in individuals' preferences, or government-related shocks like wars. Nevertheless, we will abide by the technology shock in what follows:

$$GDP_t \equiv y_t = F_t(\cdot),$$

where F is some function of production factors.

In order to measure z_t, we will take inspiration from the growth accounting technique developed by Solow. In his framework, there are two inputs, capital and labor, and a total productivity shock. Hence the previous expression takes the form:

$$y_t = F_t(K_t, n_t) = z_t F(K_t, n_t).$$

The issue that we have to address is what z is or more precisely, what the counterpart in the data to the theoretical variable $\frac{z_{t+1}}{z_t} - 1$ (the "Solow residual") is. To this effect, we will assume that time is continuous, and differentiate the production function[1]:

$$dy_t = F(K_t, n_t)\, dz_t + z_t F_K(K_t, n_t)\, dK_t + z_t F_n(K_t, n_t)\, dn_t.$$

We multiply and divide through by each component on the right hand side, so as to have percentage changes:

$$dy_t = \underbrace{z_t F(K_t, n_t)}_{y_t} \frac{dz_t}{z_t} + z_t F_K(K_t, n_t)\, K_t \frac{dK_t}{K_t} + z_t F_n(K_t, n_t)\, n_t \frac{dn_t}{n_t}.$$

Next we divide both side by total output y_t:

$$\frac{dy_t}{y_t} = \frac{dz_t}{z_t} + \frac{z_t F_K(K_t, n_t)\, K_t}{y_t} \frac{dK_t}{K_t} + \frac{z_t F_n(K_t, n_t)\, n_t}{y_t} \frac{dn_t}{n_t}. \tag{10.1}$$

With this at hand, the next task is to find the data counterparts of the fractions $\frac{z_t F_K(K_t, n_t) K_t}{y_t}$ and $\frac{z_t F_n(K_t, n_t) n_t}{y_t}$ involved in (10.1). To do this, we need to introduce two additional assumptions:

- Assumption 1: The production technology exhibits constant returns to scale.
- Assumption 2: Markets operate under perfect competition.

Assumption 1 allows for an application of the Euler Theorem:

$$F_K(K_t, n_t)\, K_t + F_n(K_t, n_t)\, n_t = F(K_t, n_t).$$

Hence each of the fractions $\frac{z_t F_K(K_t, n_t) K_t}{y_t}$ and $\frac{z_t F_n(K_t, n_t) n_t}{y_t}$ are just shares of output attributed to capital and labor respectively.

[1] We abuse the notation here a little bit. We use the time notation of a discrete process with the notation of a differential, which requires continuity (which we assume).

Assumption 2 provides the data counterpart for the derivatives F_K and F_n. Perfect competition implies that

$$\begin{aligned} R_t &= z_t F_K (K_t, n_t) \\ w_t &= z_t F_n (K_t, n_t) . \end{aligned}$$

These factor remunerations can be measured from data. Replacing in expression (10.1),

$$\frac{dy_t}{y_t} = \frac{dz_t}{z_t} + \underbrace{\frac{R_t K_t}{y_t}}_{\text{capital's share in income}} \frac{dK_t}{K_t} + \underbrace{\frac{w_t n_t}{y_t}}_{\text{labor's share in income}} \frac{dn_t}{n_t}. \tag{10.2}$$

Even though we have pinned down our empirical unknowns, measuring these is still a difficult task. Some payments are not easily classified as labor or capital income; the treatment of government and foreign trade is unclear from this expression. For further discussion on this, see Cooley (1995).

Notwithstanding the methodological difficulties, everything in expression (10.2) can be directly found in the data, except for the Solow residual dz_t/z_t, which must be solved for. This can be easily done using the following equation, which follows from (10.2) and where α_t denotes the share of capital in income at date t:

$$\frac{dz_t}{z_t} = \frac{dy_t}{y_t} - \alpha_t \frac{dK_t}{K_t} - (1 - \alpha_t) \frac{dn_t}{n_t}.$$

Also, let us fix $\alpha_t = \alpha$. A sensible value for the US (derived from data) is $\alpha \approx .3$.

Let $Z \equiv dz_t/z_t$. Then given the sequence $\{Z_t\}_{t=1950}^{1990}$, we could fit a process to this data, such as AR(1):

$$Z_{t+1} = \rho Z_t + \varepsilon_{t+1},$$

where the data show that $\widetilde{\rho} \approx .95$.

Some critiques of this approach

1. z may not be technology, but just poor measurement (Jorgenson-Griliches argument).
2. z exhibits a high variation - then what are these shocks? It should be possible to identify them. Furthermore, what is the meaning of a "negative" technological shock? Can technology somehow worsen from one period to the other?
3. The story of stochastic productivity shocks may be acceptable on an industry, or firm level. But the notion of aggregate technological shocks seems more dubious. An aggregation argument of individual, independent shocks cannot work either, since by the law of large numbers this should lead to no variation at all. Some kind of aggregate component is needed (correlation of individual shocks is tantamount to an aggregate effect).

Comments on measurement

Could the variation in z be just the product of measurement errors? It is clearly true that the variables from which the facts are observed are subject to these types of errors. In particular, the following are some of the sources of inaccuracy in measurement of some of the following variables:

(i) Total output (y):

- Quality improvements (especially in services and in government).
- Output that is not measured:
  - Learning
  - Human capital accumulation
  - Research and development.

(ii) Physical capital (K):

- Scrapping is not observed.

  In the national accounts, K is measured indirectly. Data on investment is available; hence this is used to update the registered level of capital using the accumulation equation:

$$K' = (1 - \delta) K + i.$$

- Utilization rates are not known.

(iii) Labor input into production (n):

- There is little information on the phenomenon known as "labor hoarding": personnel that is kept at their posts doing unproductive tasks.

10.3.3 Real business cycle models: brief cookbook procedure

The purpose of this section is to lay down the basic steps of the real business cycle research methodology. Using an example we will illustrate one of the most crucial steps: the calibration.

Steps to follow in real business cycle research:

1. Specify a model, including functional forms and parameters.

2. Pick parameters through calibration.

3. Solve the model numerically.

   Most often, this will be done using linearization methods. Recall that in order to do this, given an AR(1) process for the stochastic shock:

$$z' = \rho z + \varepsilon,$$

   the policy rule guesses were linear in the state variables $(K,\, z)$:

$$\begin{aligned} K' &= a_K + b_K K + c_K z \\ n &= a_n + b_n K + c_n z. \end{aligned}$$

   The task is to solve for the parameters $a_K,\, a_n,\, b_K,\, b_n,\, c_K,\, c_n$.

4. Simulate the model and analyze the outcome.

   A random number generator is used to simulate a realization of the stochastic shock. This gives rise to a time series in each of the variables. These series are the researcher's "data set". Sample moments of the variables (in general, second moments) are computed and compared to actual data.

In what follows, we will illustrate the calibration of a real business cycle model using an example. We will assume that the stochastic shock to total factor productivity follows an AR(1) process; the statistics $\widetilde{\rho}$ and $\widetilde{\sigma}^2$ need to be computed from the de-trended (HP-filtered) data.

We will assume that preferences are represented by the utility function:

$$u(c,\, l) = \frac{\left(c^{1-\theta} l^{\theta}\right)^{1-\sigma} - 1}{1-\sigma}.$$

The economy is populated by a number of identical households. Each household derives utility from the consumption of goods and leisure of its representative member. The size of household's population grows at rate η. The centralized formulation of the utility maximization problem is:

$$\max_{\{c_t,\, l_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t (1+\eta)^t u(c_t,\, l_t) \right\}. \tag{10.3}$$

The central planner faces an aggregate resource constraint:

$$C_t + X_t = A(1+\gamma)^{t(1-\alpha)} K_t^{\alpha} N_t^{1-\alpha},$$

where C_t (consumption), X_t (investment), K_t (capital), N_t (labor) denote aggregate variables. Production technology is subject to a labor-augmenting (deterministic) change process with growth rate γ.

Let P_t denote the population size at t (that grows at rate η), and divide the resource constraint by this population size:

$$\begin{aligned} \frac{C_t}{P_t} + \frac{X_t}{P_t} &= A\left(1+\gamma\right)^{t(1-\alpha)} \left(\frac{K_t}{P_t}\right)^{\alpha} \left(\frac{N_t}{P_t}\right)^{1-\alpha} \\ c_t + x_t &= A\left(1+\gamma\right)^{t(1-\alpha)} k_t^{\alpha} n_t^{1-\alpha}, \end{aligned} \tag{10.4}$$

where small-size letters denote per-capita variables. In addition, individuals' time endowment is limited, so:

$$l_t + n_t = 1. \tag{10.5}$$

The accumulation equation for capital is the usual:

$$K_{t+1} = \left(1-\delta\right) K_t + X_t.$$

Dividing through by population at t, to obtain per capita terms:

$$\begin{aligned} \left(1+\eta\right) \frac{K_{t+1}}{P_t\left(1+\eta\right)} &= \left(1-\delta\right) \frac{K_t}{P_t} + \frac{X_t}{P_t} \\ \left(1+\eta\right) k_{t+1} &= \left(1-\delta\right) k_t + x_t. \end{aligned} \tag{10.6}$$

Equations (10.3) - (10.6) constitute our problem. In order to solve them, we first transform the growth model into a stationary one. Using our previous knowledge that in this framework all variables grow at rate γ, we define the de-trended variables:

$$\widetilde{c}_t = \frac{c_t}{\left(1+\gamma\right)^t},\ \widetilde{x}_t = \frac{x_t}{\left(1+\gamma\right)^t},\ \widetilde{k}_t = \frac{k_t}{\left(1+\gamma\right)^t}.$$

We specify $\sigma = 1$ in the utility function, which leads to logarithmic utility function. Notice that

$$\log c = \log \widetilde{c} + \log\left(1+\gamma\right)^{-t},$$

but the term $\log\left(1+\gamma\right)^{-t}$ does not depend on choice variables, and hence it is irrelevant for choosing a utility maximizing consumption-leisure sequence. We ignore this term, and thus the transformed problem is:

$$\begin{aligned} &\max_{\{c_t,\, l_t,\, k_{t+1}\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t \left(1+\eta\right)^t \left[\log \widetilde{c}_t + \frac{\theta}{1-\theta} \log l_t\right]\right\} \\ &\text{s.t. } \widetilde{c}_t + \left(1+\gamma\right)\left(1+\eta\right)\widetilde{k}_{t+1} = A\widetilde{k}_t^{\alpha}\left(1-l_t\right)^{1-\alpha} + \left(1-\delta\right)\widetilde{k}_t. \end{aligned}$$

We need to pick values for the parameters involved. We begin with the ones that are immediately available:

- $\alpha = .4$ (capital share of output - constant)[2]

[2] Before we had $\alpha = .3$. Here we use a different value. The value of α is usually estimated to be around 1/3.

- $\gamma = .02$ (average long run growth rate)

- $\eta = .01$ (average long run population growth rate)

- A is a scaling factor. It is irrelevant.

- δ can be found in the following way. In the steady state of the transformed model (i.e. on the balanced growth path), we have that

$$\widetilde{k}_{t+1} = \widetilde{k}_t = \widetilde{k}^*.$$

Recall the capital accumulation equation:

$$(1+\gamma)(1+\eta)\widetilde{k}^* = (1-\delta)\widetilde{k}^* + \widetilde{x}^*.$$

Dividing both sides by $\widetilde{k}^*$ yields:

$$\begin{aligned}(1+\gamma)(1+\eta) &= 1-\delta+\frac{\widetilde{x}^*}{\widetilde{k}^*} \\ &= 1-\delta+\frac{X^*}{K^*}.\end{aligned}$$

In this equation, γ, η, and the ratio of investment to capital stock, $\dfrac{X^*}{K^*}$, are known. From the data,

$$\frac{X^{US}}{K^{US}} = .076.$$

Hence δ can be solved for: $\delta = .0458$.

Next we look at the parameters in the utility function: β and θ. For these we need to take first order conditions of the problem. Assuming a balanced growth path ($\widetilde{c}_t = \widetilde{c}_{t+1}$) we differentiate with respect to next period's capital $\widetilde{k}_{t+1}$:

$$1+\gamma = \beta\left[\alpha A\widetilde{k}_t^{\alpha-1}(1-l_t)^{1-\alpha}+1-\delta\right].$$

We can observe that

$$\alpha A\widetilde{k}_t^{\alpha-1}(1-l_t)^{1-\alpha} = \alpha\frac{\widetilde{y}_t}{\widetilde{k}_t} = \alpha\frac{Y_t}{K_t},$$

where $\frac{Y_t}{K_t}$ is available from actual data (annual output data):

$$\frac{Y_t^{US}}{K_t^{US}} \approx .3012.$$

With this, β can be solved for. The result is $\beta = .94912$.

The parameter θ is more controversial. We need the remaining first order conditions, which are:

$$\begin{aligned} \widetilde{c}_t &: \quad \frac{1}{\widetilde{c}_t} = \lambda_t \\ l_t &: \quad \frac{\theta}{1-\theta}\frac{1}{l_t} = \lambda_t (1-\alpha) A \widetilde{k}_t^{\alpha} (1-l_t)^{-\alpha}. \end{aligned}$$

Then the Euler equation determining labor supply is:

$$\begin{aligned} \frac{\theta}{1-\theta}\frac{1-l_t}{l_t} &= (1-\alpha)\frac{\widetilde{y}_t}{\widetilde{c}_t} \\ &= (1-\alpha)\frac{Y_t}{C_t}. \end{aligned}$$

Let us first look at $\frac{Y_t}{C_t}$. We have that

$$\frac{Y_t}{C_t} = \frac{Y_t}{K_t}\frac{K_t}{C_t}$$

and

$$\begin{aligned} & X_t + C_t = Y_t \\ \Rightarrow\ & \frac{X_t}{K_t} + \frac{C_t}{K_t} = \frac{Y_t}{K_t} \\ \Rightarrow\ & \frac{C_t}{K_t} = \frac{Y_t}{K_t} - \frac{X_t}{K_t}. \end{aligned}$$

Since we know the values of $\frac{Y_t}{K_t}$ and $\frac{X_t}{K_t}$ from actual data we can find $\frac{Y_t}{C_t}$.

The next issue is what a reasonable estimate of l_t is. In this case, we must use knowledge from microeconomic studies. We can see that out of the total 24 hours daily endowment, 8 hours are used for sleeping, 8 for work, and the remaining 8 for leisure. Then we can use $l_t \approx 2/3$. Using this, we can solve for θ, which yields $\theta = .61612$.

The methodology used is controversial from a microeconomic point of view, due to the response to wage changes that it implies. In the decentralized problem, the first order condition for leisure is:

$$\frac{\theta}{1-\theta}\frac{1}{1-n_t} = \frac{1}{\widetilde{c}_t}\widetilde{w}_t$$

$$\Rightarrow (1-n_t) w_t = \frac{\theta}{1-\theta} c_t.$$

We want to examine the change in labor supply n_t arising from a change in the wage w_t. There are several ways to measure this elasticity, but let us use:

$$\left.\frac{d\log n_t}{d\log w_t}\right|_{c_t \text{ constant}} \equiv \xi.$$

This is called the "λ-constant" labor supply elasticity. From the first order conditions, we have

$$\begin{aligned}
n_t &= 1 - \frac{\theta}{1-\theta}\frac{c_t}{w_t} \\
\log n_t &= \log\left(1 - \frac{\theta}{1-\theta}\frac{c_t}{w_t}\right) \\
\Rightarrow \xi &= \frac{\dfrac{\theta}{1-\theta}\dfrac{c_t}{w_t}}{1 - \dfrac{\theta}{1-\theta}\dfrac{c_t}{w_t}} \\
&= \frac{1-n_t}{n_t}.
\end{aligned}$$

So inserting the value of $n_t = 1/3$ used above yields $\xi = 2$.

But this result is wildly contradicted by data. Labor economists have found that this elasticity is approximately zero - hours worked are very little sensitive to wage changes. This is a serious drawback for the calibration used: parameter values are not consistent with microeconomic data.

10.3.4 Real business cycle in action: A model example

The purpose of this section is to briefly give an example of an actual model. This is a model developed by Hansen (1985) and Rogerson (1988) to introduce equilibrium unemployment, the lack of which is a shortfall in the usual real business cycle models.

Hansen and Rogerson model an economy which is populated by many (a measure one of) agents and which contains some sort of fixed cost to working. As a consequence, the employment decision is a discrete variable: n_t can only take the values $\overline{n}$ or 0. Hence, leisure can only be either $1 - \overline{n}(\equiv \overline{l})$ or 1.

The main problem with this assumption is that a competitive equilibrium may not exist. In order to overcome this difficulty, the authors introduce an employment lottery, whereby individuals become employed in a full-time job with probability $1 - \mu$, and unemployed otherwise. This plays the role of "convexifying" the decision making environment and leads to the applicability of the usual existence and welfare theorems.

The central planner maximizes the following function:

$$E[(1-\mu)\,(\log c_e + \frac{\theta}{1-\theta}\log \overline{l}) + \mu(\log c_u + \frac{\theta}{1-\theta}\underbrace{\log 1}_{0})],$$

where the expectation is taken across agents - hence it is in fact a weighted sum of utilities. The outcome from solving the maximization problem will be a point in the Pareto frontier of the economy (recall the Negishi characterization). The term in the brackets is an individual agent's expected utility for a given period. c_e is the consumption level that the planner assigns to working individuals, and c_u to the unemployed.

The resource constraint that the planner faces is:

$$(1-\mu)\, c_e + \mu c_u + \bar{i} = zK^{\alpha}\left[(1-\mu)\,\overline{n}\right]^{1-\alpha},$$

where the law of large numbers was used to assert that μ is the fraction of the population that will be unemployed in a given period. The choice variables for the planner are the consumption levels c_e and c_u, the probability of unemployment μ, and the aggregate investment $\bar{i}$.

Ignoring the law of motion for capital, take $\bar{i}$ as given and solve the resulting "sub-optimization" problem to find c_e, c_u and μ. The first order conditions are:

$$\begin{aligned} c_e &: \quad \frac{1-\mu}{c_e} = \lambda\,(1-\mu) \\ c_u &: \quad \frac{\mu}{c_u} = \lambda\mu \\ \\ \Rightarrow & \qquad c_e = c_u. \end{aligned}$$

This is a complete markets result: complete risk sharing. The conclusion is that in this model, the employed individuals are the unlucky ones. We can use the result $c_e = c_u = c$ to reformulate the problem for μ:

$$\begin{aligned} &\max_{\mu}\left\{\log c + (1-\mu)\,\frac{\theta}{1-\theta}\log\bar{l}\right\} \\ &s.t.\ c + \bar{i} = zK^{\alpha}\left[(1-\mu)\,\overline{n}\right]^{1-\alpha}. \end{aligned}$$

Then $(1-\mu)$ can be viewed as total hours worked ($\overline{n}$ is just a normalization). So the result is that we have modified preferences, that evolved from being logarithmic in leisure to being actually linear in leisure (notice that since $\bar{l} < 1$, an increase in μ leads to an increase in utility). As a result of this linearity, labor supply will be very elastic in the aggregate, in spite of being totally inelastic at the individual level.

10.4 Other real business cycle models

10.4.1 Investment-specific technology shocks

here GHK

10.4.2 Fiscal policy and government spending shocks

here aiyagari

10.4.3 Monopolistic competition and markup shocks

here hornstein

10.4.4 Preference shocks

here gali model

10.4.5 Models with adjustment costs in technology and preferences

here habits and tobin's q

10.5 Sunspot models of the business cycle

The main advocates of these models are Farmer and Azariadis. The main characteristics of their approach are:

- There are no fundamental (or intrinsic) shocks.

- There is a competitive equilibrium. Aggregate variables (c, y) will fluctuate randomly.

***Q*:** Could this happen in an environment where the equilibrium is Pareto optimal?

***A*:** No (assuming strict risk aversion). If consumption fluctuates but could be constant, this cannot be an optimal result.

These models show either or both:

- Distortions
- Externalities

In real-business-cycles-like macroeconomic models, it is possible to prove that sunspot equilibria exist whenever the equilibrium is indeterminate. Recall the second-order difference equations arising from the linearization of the neoclassical growth model. If the two roots were $|\lambda_1| < 1$, $|\lambda_2| > 1$, then we were able to rule out an exploding path. However, if both λ_1 and λ_2 resulted in values smaller than 1 in absolute value, we ran into "indeterminacy" - several convergent paths were equally admissible as a solution.

Then sunspot equilibria are the result of a randomization among the convergent paths. The procedure in this research area is to build real-business-cycles-like models and play with the parameters until a situation with $|\lambda_1| < 1$ and $|\lambda_2| < 1$ is reached. This is not an easy task, and demands a high level of distortions and externalities from the model.

10.6 References

Burns, Arthur F., and Wesley C. Mitchell, "Measuring Business Cycles," *National Bureau of Economic Research*, 1946.

Cooley, Thomas F., “Frontiers of Business Cycles Research”, Princeton University Press, 1995, Chapter 1.

Hansen, Gary D., “Indivisible Labor and the Business Cycle,” *Journal of Monetary Economics*, 16(1985), 309-27.

Koopmans, Tjalling C., “Measurement Without Theory,” *Review of Economics and Statistics*, 3(1947), 161-172.

Kydland, Finn E. and Edward C. Prescott, “Time to Build and Aggregate Fluctuations,” *Econometrica*, 50(1982), 1345-1370.

Long, John B. and Charles I. Plosser, “Real Business Cycles,” *Journal of Political Economy*, 91(1983), 39-69.

Lucas, Robert E., Jr., “Econometric Policy Evaluation: A Critique,” *Journal of Monetary Economics*, 1(1976), 19-46.

Lucas, Robert E., Jr., “Understanding Business Cycles,” *Journal of Monetary Economics*, 5(1977), 7-29.

Prescott, Edward C., “Theory ahead of business cycles measurement”, textitMinneapolis Federal Reserve Quarterly Report, 1986.

Rogerson, Richard, “Indivisible labor, lotteries and equilibrium,” *Journal of Monetary Economics*, 21(1988), 3-16.

Chapter 11

Frictional labor markets

This chapter looks at a number of dimensions in which labor markets face frictions. First, in all developed economies taxes on labor earnings constitute a major source of government revenues and therefore we need to look at how such a tax acts on labor supply from a macroeconomic perspective. Second, labor supply is often made subject to an "indivisibility". That is, even though the previous analysis treated hours as a continuous variable, at least from the perspective of a single household, it is not necessarily realistic to describe households as being able to vary the hours worked, and hence earnings, continuously; in fact, our main means for understanding labor supply was through a first-order condition where hours are assumed to be fully subject to the household's choice. Thus, we will now analyze the polar opposite case, i.e., one where hours worked are either 0 or some preset number of hours. We will thus first show how a market with indivisibilities might work and conclude that the answer depends crucially on the market structure for consumption insurance across households. Thus, the indivisibility model will typically have not all agents working at the same time, though in the absence of frictions other than the indivisibility itself the market outcome will be efficient. An agent not working will thus be regarded as "out of the labor force".

Third, and most importantly, this chapter will analyze unemployment: in practice, many workers would like to work at the prevailing market work (that is relevant to their skills) but cannot find a job. The main tool of the chapter is the Diamond-Mortensen-Pissarides (DMP) model of search/matching frictions. This model takes a specific view on why unemployment exists in the first place, and it has become a workhorse for macroeconomists. This model has been developed further and generalized in a number of directions since it was first constructed, but it is of course not the only relevant model of unemployment. A closely related model is the "island search model" due to Lucas and Prescott (197xyz), which emphasizes that unemployment, though literally making workers sometimes not able to work even though they would like to, does not have to be associated to economic inefficiency; in their setting, exogenous shocks force reallocation of workers across activities and markets always produce Pareto-optimal outcomes, at least in the presence of insurance markets for consumption. This model will be briefly discussed and compared to the DMP model. Yet another model is the efficiency-wage model, which makes assumptions implying firms pay workers more than their short-run productivity and hence, at such wages, there is an excess supply of labor and, hence, unemployment. Efficiency-wage models also rely on frictions (either in search or matching), and they will not be discussed in much detail in this text. Neither will models where unemployment is due to union power; clearly, if unions are mostly concerned for the already employed workers and their wages and not for the unemployed, their wage demands will too high to admit full employment. The omission of these models in the text is mainly made so as to contain the scope of the text, the DMP model being the most frequently used model in practice. It is clear that other models of unemployment capture important phenomena; for example, many argue that unions have played a central role in European labor markets at least during significant periods of time.

Finally, the labor market is a market where there is striking heterogeneity across households: not only are some workers employed and other unemployed, but wages per hours worked vary a lot across workers, and hours worked per year among the non-unemployed also display rather wide dispersion. What factors explain these phenomena? Determinants of the observed wage inequality include educational differences across workers, other skill differences, compensation for differences in job amenities ("compensating wage differentials"), and simply luck. Arguably these factors are all quantitatively important factors for people's lives, but the degree to which they are determined by frictions is less clear. Therefore, the discussion of inequality in labor markets will be postponed until Chapter 14, where inequality is discussed more broadly.

11.1 Taxes in a labor market without frictions

Going back to the determinants of long-run economic performance, it is clear that countries where citizens work more will, everything else equal, have a high relative GDP per capita. The differences in hours worked across countries are not negligible, even within the developed world. Indeed, Prescott (2004) argues precisely that an important reason why an average European country has seen its GDP per capita fall relative to the U.S. over the postwar period is a relative lowering of the labor input into production. Prescott further argues that higher taxes on work in Europe accounts for the lion's share of this difference. These arguments can be explained and evaluated based on the material covered in this chapter. To assess and compare welfare across countries, of course, one also needs to take into account how leisure affects utility; this issue was discussed in the context of the Jones and Klenow (2011) paper above. The chief purpose of the section is thus to examine how a benchmark economy—one with frictionless markets—deals with the determination of labor input in the presence of taxes on earnings..

11.1.1 Proportional taxes on labor earnings

Although hours worked in the frictionless model without capital, i.e., where production is linear in labor, are determined by preferences only, a tax rate on labor income could influence labor supply. How can that be, given that such a tax could be viewed as a decrease in the wage, which we know will not change the amount of hours worked? The answer is that labor supply may or may not change; the key here is how the government uses the tax proceeds. If the government simply wastes the tax revenue (or spends them on some good that does not influence either the marginal utility of consumption or leisure), the labor income tax will indeed have no effect. But if the government rebates all (or part) of the revenue back to the consumer in a lump-sum fashion, a raised tax will lower labor supply. Intuitively, now there is no (or a smaller) income effect in equilibrium of the lower net-of-tax wage, since there will be compensating income through the transfer, and so the substitution effect will dominate. Formally, the first-order condition will read $v(l)w(1-\tau) = cv'(l)$, where τ is the tax rate. The budget reads $c = w(1-l)(1-\tau) + rk + T$, where $T = w(1-l)\tau$. So in equilibrium the budget will read $c = w(1-l)+rk$ and therefore the equation determining hours worked (for a given wage, rental rate, and capital level) is $v(l)(1-\tau) = (1-l+rk/w)v'(l)$.

11.1.2 Data

Prescott's (19xyz) paper ... [HERE GO THROUGH THE NUMBERS AND PROVIDE SOME CRITICAL COMMENTARY.]

Also perhaps comment on progressive taxation here.

11.2 Indivisible labor

Next we consider a model where labor supply is indivisible and binary, i.e., that it can only take on the values 0 or 1. The reason is that there may be an important fixed cost in working. Thus, we can think about the labor-supply decision as one of participating in the labor market or not. First consider a static economy and suppose that there is a continuum of mass 1 of consumers who are all identical and value consumption and leisure $u(c,l)$ as described above. Moreover, there is a production technology; to make the main point as starkly as possible, let us assume first that it is linear in labor: $y = A\mu$, where μ is the fraction of the population working. Consider a competitive market where firms hire workers, so that—at least if any workers are hired—the wage would have to equal the marginal product: $w = A$. Workers then choose whether to work or not, i.e., they would compare $u(w,0)$ to $u(0,1)$. Supposing that we use a utility function of the form $\log c - B(1-l)^\nu/\nu$ where $\nu \geq 1$, it is clear that work would be preferred at any positive wage: the disutility of working is bounded, but the utility of zero consumption is unboundedly negative. Thus the outcome in competitive equilibrium is that all consumers work and obtain utility $\log A - B/\nu$. Is this outcome, however, the best possible? Suppose one worker were to not work, thus gaining B/ν utils, resulting in a total resource loss of A units of consumption. If all consumers shared this loss equally, clearly we would not be looking at a Pareto improvement—since all the other workers would obtain a lower level of consumption—but average utility might go up. The utility loss in consumption if borne equally is $Au_c(A,0) = 1$. So if the utility gain by the non-worker is higher than this, i.e., $B/\nu > 1$, this change would increase average utility in the population. Suppose the parameter values indeed satisfy this inequality. Wouldn't there be a way of engineering a Pareto-improving trade?

The answer is yes and it relies on the use of "lotteries". Suppose, prior to production and consumption, people met and realized about the calculation above and agreed to implement a random allocation of consumption and working hours such that all agents are treated equally ex ante but different ex post. Then what kind of lottery would they agree upon? Let $(c_p, 0)$ and $(c_{np}, 1)$ be the allocation of consumption and leisure for the two ex-post groups, where the subindex p refers to "participants" and np to "non-participants". Thus, within each of the two groups, all agents are treated equally (there is no gain from randomization). The fraction of p agents, a choice variable for the group as a whole, is denoted μ and also is the probability with which a consumer is assigned to work. The resource constraint thus reads$\mu c_p + (1-\mu)c_{np} = A\mu$. The maximization problem of the group as a whole is

$$\max_{\mu, c_p, c_{np}} \mu \log c_p + (1-\mu)\left(\log c_{np} - B/\nu\right) \quad \text{s.t.} \quad \mu c_p + (1-\mu)c_{np} = A\mu \quad \text{and} \quad \mu \leq 1.$$

Clearly, the consumption choices imply $c_p = c_{np} = A\mu$: the separability of the utility function implies that setting the marginal utility of consumption equal for all agents—which is optimal since consumers are risk-averse—is optimal. Hence the maximization problem reduces to choosing μ so as to maximize $\log(A\mu) - (1-\mu)B/\nu$. The solution, if interior, is to set $\mu = \nu/B$. If this value is one or greater, the optimal choice is for all workers to work. Clearly, this condition is exactly the condition we arrived at above in reallocating one agent away from working ex post. With the lottery allocation, a Pareto improvement has been made possible if $\nu/B < 1$. One might worry that this finding depends on having assumed that there is no other income than labor income, but it is not. It is straightforward to introduce capital, or any other, income, and the lottery allocation will still improve on the competitive equilibrium with the indivisibility.

Is it possible to arrive at the lottery outcome in a competitive market context? Yes, the market could have trade in lotteries: the commodity space would be a lottery space. Different kinds of lotteries could then be priced and consumers would choose among them; the outcome, shown by Rogerson (1988), would deliver the allocation just derived.

A striking feature that appeared in deriving the result is that the "planner utility", after the optimal choice of consumption—which is set equal for all agents—becomes $\log c - \mu B/\nu$, with consumption given as $c = A\mu$. Now think of μ as employment: total labor input, which is equivalent here to the total number of people working. Thus we can instead write $\log c - (B/\nu)n$, with $c = An$. The "reduced-form" utility function now appears as linear in labor (or leisure): its Frisch elasticity of labor supply is therefore infinite, since it corresponds to the case of $\gamma = \infty$. The utility function is still in the class where income and substitution effects cancel, so in the static model just analyzed changes in productivity would have no effect on employment. In an intertemporal context, however, provided that there is a technology for saving and that the decreasing returns are not too strong, a model with indivisible labor and lotteries would imply very strong swings in labor supply in response to productivity. Thus, the high elasticity is arrived at without having to assume anything particular about preferences: the key is the indivisibility and the availability of lotteries as a way of dealing with the indivisibilities.

An employment lottery is an abstract concept; indeed in the static model it is hard to come up with an alternative way to improve utility. In an intertemporal context, however, lotteries are not necessary for improving utility compared to the allocation where everyone works. The idea is that it is possible to support a lottery-like outcome with borrowing and lending among agents. Suppose a set of identical agents meet at time zero and trade in a competitive market for labor (with an indivisibility) and borrowing and lending, at a competitively determined interest rate. Then the outcome will be one where people have the same consumption, which is smoothed over time, and what differs across people is when they work. An agent who does not work in the first period will work at future times and therefore can borrow against the future labor income to support first-period consumption. Such a model is described in Krusell, Mukoyama, Rogerson and Sahin (2008). The point here is that a high intertemporal elasticity of total employment will arise in such a model, and they key is not the existence of lotteries but the existence of a market for borrowing and lending. [PERHAPS ADD A SIMPLE VERSION OF THIS MODEL HERE.]

11.3 Theories of unemployment

So far all the models considered have been models of employment or hours worked without frictions. We now turn to some models of frictional unemployment. The core setup will be based on search and matching frictions. We begin with a cornerstone of this literature: the search model due to McCall xyz and Mortensen xyz.

11.3.1 Search theory

Consider an economy populated by ex-ante equal, risk-neutral, infinitely lived individuals who discount the future at rate r. Unemployed agents receive job offers at the instantaneous rate λ_u. Conditionally on receiving an offer, the wage is drawn from a well-behaved distribution function $F(w)$ with upper support $w^{\max}$. Draws are i.i.d. over time and across agents. If a job offer w is accepted, the worker is paid a wage w until the job is exogenously destroyed. Separations occur at rate σ. While unemployed, the worker receives a utility flow b which includes unemployment benefits and a value of leisure and home production, net of search costs. Thus, we have the Bellman equations

$$rW(w) = w - \sigma\left[W(w) - U\right] \tag{11.1}$$

$$rU = b + \lambda_u \int_{w^*}^{w^{\max}} \left[W(w) - U\right] dF(w), \tag{11.2}$$

where $rW(w)$ is the flow (per period) value of employment at wage w, and rU is the flow value of unemployment.

[1]

In writing the latter, we have used the fact that the optimal search behavior of the worker is a reservation-wage strategy: the unemployed worker accepts all wage offers w above $w^* = rU$, at a capital gain $W(w) - U$. Solving equation (11.1) for $W(w)$ and substituting in (11.2) yields the reservation-wage equation

$$w^* = b + \frac{\lambda_u}{r+\sigma} \int_{w^*}^{w^{\max}} [w - w^*] \, dF(w) .$$

Without loss of generality, let $b = \rho \bar{w}$, where $\bar{w} = \mathbb{E}[w | w \geq w^*]$. Then,

$$\begin{aligned} w^* &= \rho \bar{w} + \frac{\lambda_u [1 - F(w^*)]}{r+\sigma} \int_{w^*}^{w^{\max}} [w - w^*] \frac{dF(w)}{1 - F(w^*)} \\ &= \rho \bar{w} + \frac{\lambda_u^*}{r+\sigma} [\bar{w} - w^*] , \end{aligned} \tag{11.3}$$

where $\lambda_u^* \equiv \lambda_u [1 - F(w^*)]$ is the job-finding rate. Equation (11.3) relates the lowest wage paid (the reservation wage) to the average wage paid in the economy through a small set of model parameters.

If we now define the mean-min wage ratio as $Mm \equiv \bar{w}/w^*$ and rearrange terms in (11.3), we arrive at

$$Mm = \frac{\frac{\lambda_u^*}{r+\sigma} + 1}{\frac{\lambda_u^*}{r+\sigma} + \rho}. \tag{11.4}$$

The mean-min ratio Mm is our new measure of frictional wage dispersion, i.e., wage differentials entirely determined by luck in the random meeting process. This measure has one important property: it does not depend *directly* on the shape of the wage distribution F. Put differently, the theory allows predictions on the magnitude of frictional wage dispersion, measured by Mm, without requiring any information on F. The reason is that all that is relevant to know about F, i.e., its probability mass below w^*, is already contained in the job finding rate λ_u^*, which we can measure directly through labor market flows from unemployment to employment and treat as a parameter.

[1] This equation is written in flow form but can be derived from a discrete-time formulation. Take the first equation, (11.1), for illustration. Suppose that the value of having a job is constant over time from the perspective of a matched firm, and that we are looking at one period being of length Δ. During this period, there is production and wages are paid, the net amount being $(p-w)\Delta$, since p and w are measured per unit of time. At the end of the period, the match separates with probability $\sigma\Delta$ and remains intact with probability $1 - \sigma\Delta$. So it must be that $J(t) = (p-w)\Delta + (1-\sigma\Delta)e^{-r\Delta}J(t+\Delta) + \sigma\Delta e^{-r\Delta}V$. Here, $e^{-r\Delta} \equiv \delta(\Delta)$ is a discount factor; it gives a *percentage* decline in utility as a function of the length of time, $-(d\delta(\Delta)/d\Delta)/\delta(\Delta)$, which is constant and equal to r. Subtract $J(t+\Delta)e^{-r\Delta}$ on both sides and divide by Δ. That delivers $\frac{J(t)-J(t+\Delta)}{\Delta} + \frac{J(t+\Delta)(1-e^{-r\Delta})}{\Delta} = p - w - \sigma e^{-r\Delta}(J(t+\Delta) - V)$. Take limits as $\Delta \to 0$. Then the left-hand side becomes $\dot{J}(t) + rJ(t)$, the second term coming from $J(t+\Delta)$ being continuous and going to $J(t)$ and application of l'Hôpital's rule. The right-hand side gives $p - w - \sigma(J(t) - V)$. In a steady state $J(t)$ is constant and equal to a J satisfying the equation in the text.

The model's mean-min ratio can thus be written as a function of a four-parameter vector, $(r, \sigma, \rho, \lambda_u^*)$, which we can try to measure independently. Thus, looking at this relation, if we measure the discount rate r to be high (high impatience), for given estimates of σ, ρ, and λ_u^*, an increased Mm must follow. Similarly, a higher measure of the separation rate σ increases Mm (because it reduces job durations and thus decreases the value of waiting for a better job opportunity). A lower estimate of the value of non-market time ρ would also increase Mm (agents are then induced to accept worse matches). Finally, a lower measure of the contact rate λ_u^* pushes Mm up, too (because it makes the option value of search less attractive) The assumptions here Search without OJS

Search with OJS

Search at a cost

The Diamond paradox

Frictional wage dispersion

11.3.2 The Pissarides model

This is a continuous time model in which:

- all workers are risk-neutral and alike
- all firms are alike
- a match produces an output, p
- wage $w(t)$ a result of a Nash Bargaining solution
- workers meet firms with probability $\lambda_w(t)$
- firm meets worker with probability $\lambda_f(t)$
- it costs to create a vacancy
- there is free entry
- unemployed workers get $b > 0$, a measure unemployment benefit or a valuation of leisure

$$\dot{u}(t) = \underbrace{\underset{\substack{\text{exogenous}\\\text{separation}\\\text{rate}}}{\sigma} \quad (1 - u(t))}_{\substack{\text{flow}\\\text{into unemployment}}} - \underbrace{\underset{\substack{\text{endogenous flow probability}\\\text{that unemployed worker}\\\text{finds a job}}}{\lambda_w(t)} \quad u(t)}_{\substack{\text{flow}\\\text{out of unemployment}}} \tag{11.5}$$

In steady state: $\dot{u}_t = 0$

$$\Rightarrow \sigma(1 - u(t)) = \lambda_w(t)u(t) \tag{11.6}$$

Define the matching function as

$$m(t) = m(u(t), v(t)) \tag{11.7}$$

$m(t)$ exhibits constant returns to scale and is concave in both its arguments: unemployment $u(t)$ and vacancies $v(t)$. Define the market tightness as:

$$\theta = \frac{v(t)}{u(t)} \tag{11.8}$$

and the endogenous flow probability that worker finds a job as:

$$\lambda_w(t) = \frac{m(t)}{u(t)} \tag{11.9}$$

and the endogenous flow probability that the firms fills a vacancy as:

$$\lambda_f(t) = \frac{m(t)}{v(t)} \tag{11.10}$$

The CRS property is attractive. Had one chosen an IRS matching function, one would create a strong search externality as it would imply that the more people there are of any kind in this economy, the more likely that they will find a match. With CRS, the scale is obviously independent.

Let us first consider the steady state conditions and then allow for $p = p(t)$. As p is what occurs when a match takes place, if we let it fluctuate, it may be considered an element of the RBC models and one can try to calibrate the model. This has, however, been criticized for not giving good predictions.

Steady state analysis Let r be a constant, exogenous interest rate. Consider the firm's side of this economy:

Firms

$$rJ = p - w - \sigma(J - V) \tag{11.11}$$

The Bellman equation above describes an asset value for the firm to have a worker employed. J is the value of the firm with a worker and hence rJ is the flow value having a position filled. p is the flow productivity and w is the wage paid. Hence $p - w$ is what it pays to have a worker employed. $\sigma(J - V)$ shows the loss in case a change of state (from having a job filled, J, to holding a vacancy, V) occurs. The next Bellman equation,

$$rV = -c - \lambda_f(J - V) \tag{11.12}$$

describes the asset value of holding a vacancy for the firm. V is the value of the firm when holding a vacancy open, c is the cost the firm pays to post a vacancy and λ_f is the probability that the firm fills a vacancy. Analogously for the workers:

Workers

$$rW = w - \sigma(W - U) \tag{11.13}$$

$$rU = b + \lambda_w(W - U) \tag{11.14}$$

Where W is the value of employment and U is the value of being unemployed. λ_w is the flow probability of finding a job. Note that b is an exogenous parameter.

Wage determination In this model, the wage, w is determined by bargaining:

$$\begin{aligned} w &= \arg\max \left[(W(w) - U)^{\beta} \left(J(w) - V\right)^{1-\beta}) \right] && (11.15) \\ \text{FOC} &\Rightarrow \beta(J(w) - V)\frac{\partial W(w)}{\partial w} - (1-\beta)\left(W(w) - U\right)\frac{\partial J(w)}{\partial w} = 0 \end{aligned}$$

Where β is a parameter measuring the bargaining power of the workers. Noting that the individuals are risk-neutral (there is no curvature in the utility), we can define the surplus of a match as:

$$S \equiv (J(w) - V) + (W(w) - U) \geq 0 \tag{11.16}$$

and re-write the the bargaining in terms of what fractions of the surplus will go to which of the bargaining parties:

$$\begin{aligned} \text{Firms:} \qquad & S(1-\beta) = (J(w) - V) && (11.17) \\ \text{Workers:} \qquad & S\beta = (W(w) - U) && (11.18) \end{aligned}$$

Adding and subtracting the Bellmans yields:

$$\begin{aligned} rS &= r(J - V) + (W - U) = \\ & [p - w - \sigma(J - V)] - \\ & [-c - \lambda_f(J - V)] + \\ & [w - \sigma(W - U)] - \\ & [b + \lambda_w(W - U)] \\ &= p - \sigma S + c - \lambda_f S(1-\beta) - \lambda_w S\beta - b && (11.19) \end{aligned}$$

and the solution is: $S = \frac{-b+c+p}{r+\sigma+\beta\lambda_w+\lambda_f(1-\beta)}$ if $r + \sigma + \lambda_f - \beta\lambda_f + \beta\lambda_w \neq 0$

Now take

$$\begin{aligned}
rJ - rW &= p - w - \sigma(J - V) - rW, \text{ re-write:} \\
r(1-\beta)S &= p - w - \sigma(1-\beta)S - rW
\end{aligned}$$

$$\begin{aligned}
\text{but } rS &= p - \sigma S + c - \lambda_f S(1-\beta) - \lambda_w S\beta - b, \;\; \shortparallel \times (1-\beta) \text{ becomes:} \\
r(1-\beta)S &= (1-\beta)[p - \sigma S - rV - rU] \\
& \quad \text{now subtract the two} \\
&\Rightarrow \;\; 0 = p - w - \sigma(1-\beta)S - rW - ((1-\beta)[p - \sigma S - rV - rU]) \text{ and simplifying:} \\
&\Rightarrow \;\; w = \beta(p - rV) + (1-\beta)rU \qquad (11.20)
\end{aligned}$$

The above is the flow version of the wage function. It can be re-written as:

$$w - rU = \beta(p - rV - rU) \qquad (11.21)$$

Which gives us an expression of the workers' surplus payment of being employed; in other words it is the workers' fraction of the total flow surplus value. Let's make use of the free entry assumption: $V = 0$

$$\begin{aligned}
r0 &= -c - \lambda_f(J - 0) \\
&\Rightarrow \;\; J = -\frac{c}{\lambda_f}
\end{aligned}$$

But then

$$\begin{aligned}
S(1-\beta) &= (J(w) - 0) \\
S(1-\beta) &= -\frac{c}{\lambda_f} \\
&\Rightarrow \;\; S = -\frac{c}{\lambda_f(1-\beta)}
\end{aligned}$$

Set the expressions for S equal:

$$\begin{aligned}
S &= -\frac{c}{\lambda_f(1-\beta)} = \frac{-b + c + p}{r + \sigma + \beta\lambda_w + \lambda_f(1-\beta)} \\
&\Rightarrow \;\; \frac{-b + p}{r + \sigma + \beta\lambda_w} = \frac{c}{\lambda_f(1-\beta)} \qquad (11.22)
\end{aligned}$$

If we assume the matching function to be of the Cobb-Douglas form:

$$\begin{aligned}
m &= Au^{\alpha}v^{1-\alpha} \text{ and we use that } \theta = \frac{v}{u} \\
&\Rightarrow \;\; \lambda_f = A\theta^{-\alpha} \text{ and} \\
&\Rightarrow \;\; \lambda_w = A\theta^{1-\alpha}
\end{aligned}$$

And so we can solve for u and θ. Note, however, that we will not obtain a closed form solution for θ, but we can perform comparative statics analysis. We have that

$$w - rU = \beta(p - rV - rU)$$

and

$$u = \frac{\sigma}{\sigma + \lambda_w}$$

$$\text{Let } \hat{\theta} \equiv \frac{d\theta}{\theta} \tag{11.23}$$

$$\Rightarrow \hat{\theta} = \frac{r + \sigma + \beta\lambda_w}{\alpha(r+\sigma) + \beta\lambda_w}\left[\frac{p}{p-b}\hat{p} - \frac{b}{p-b}\hat{b} - \frac{\sigma}{r+\sigma+\beta\lambda_w}\hat{\sigma} - \hat{c}\right] \tag{11.24}$$

and

$$\hat{u} = -(1-u)(1-\alpha)\hat{\theta} \tag{11.25}$$

And so

$$\begin{aligned} \frac{\hat{\theta}}{\hat{p}} &= \frac{d\theta}{\theta}/\frac{dp}{p} = \frac{r+\sigma+\beta\lambda_w}{\alpha(r+\sigma)+\beta\lambda_w}\frac{p}{p-b} \\ &= \frac{r+\sigma+\beta\lambda_w}{\alpha(r+\sigma)+\beta\lambda_w}\frac{p/p}{(p-b)/p} \\ &= \frac{r+\sigma+\beta\lambda_w}{\alpha(r+\sigma)+\beta\lambda_w}\frac{1}{1-\frac{b}{p}} \end{aligned} \tag{11.26}$$

When studying the data, one sees that θ is strongly pro-cyclical, but when Shimer calibrated the model he found that the market tightness is only mildly pro-cyclical. Shimer concluded that for it to perform well $\frac{b}{p} \approx 1$, however the U.S. data suggests this should be $\frac{b}{p} \approx 0.4$. Big fluctuations in $\hat{u}$ should be interpreted as fluctuations in $\hat{p}$, when calibrated, however, this failed. If b is high than the workers ought to get most of the surplus: $w \sim p \sim b$. A small change in p should lead to large percentage increase in the firms' profits and that should lead to a drop in unemployment. This is however not reasonable with the U.S. data. With $\frac{b}{p} \approx 0.4$ the fluctuations are very small in u. Hence, the basic problem with the model is that an increase in p predicts too large increase in the wages. The workers get too much of p and so the firms do not flow fast enough. If one however models rigid wages, $w = \bar{w}$, then there exists enough incentive for firms to flow in. Hence with real wage rigidity, the model performs better.

Dynamics/out of steady state analysis What are the state and jump variables in this model? Unemployment u is the state variable, while vacancies, v is the jump variable. θ takes on the property of v and is also a jump variable.

11.3.3 The European unemployment dilemma

Ljungqvist and Sargent (1998) introduce a turbulence shock in the level of wages on the individual level is studied. The authors assume there are different levels of human capital: h_g and h_b. They argue that the outside option in Europe is so attractive, due to that the level of the benefits is about 80 percent of the wages, so that workers choose to stay unemployed. Hence the workers in Europe are not as keen on searching as in the U.S..

[Here there followed graphs of unemployment in Europe and U.S. from 1950-2000 and a graph of $\frac{w_s}{w_u}$ over time. We also briefly discussed the paper by Katz and Murphy]

[MENTION HORNSTEIN-KRUSELL-VIOLANTE]

A discussion on the 1970's oil shocks followed. The question is whether the oil shock were an important factor or rather if the recession was due to a productivity slowdown.

11.3.4 Rogerson

In this paper a different view on deterioration of the labor market is presented. First, it is commonly argued that the deterioration of employment in Europe began in the 1970's. But if one examines the European employment relative to the U.S. one notes that the quantitative change came earlier and was bigger. Rogerson argues that if one instead focuses on labor input, one obtains a different characterization of this deterioration. In particular, the deterioration of the European labor market outcomes relative to the U.S. began earlier, in the 1950's. Hence perhaps there might be a forgotten factor in the European unemployment dilemma. Rogerson argues this is due to a structural change in how the sectors have grown and developed. Services, industry and agriculture have developed differently in Europe and the U.S.. Rogerson suggests that Europe failed to transform away from a large agricultural sector as fast as the U.S. The U.S. on the other hand has a large services sector. The reason for why this has not been examined earlier is that measuring of the output in the services sector is far harder then in other industries.

11.4 Inequality in the labor market

What explains differences in $\frac{w_s}{w_u}$? Suppose we have labor supply that is not homogenous, but rather we have L_s and L_u, the labor input of skilled and unskilled workers. Further suppose we have perfect competition (i.e. this is no longer the search theoretical framework). Characterize the labor demand by:

$$Y = F(K, L_s, L_u) = K^\alpha G(L_s, L_u)^{1-\alpha} \tag{11.27}$$

Which is CES and exhibits CRS. Since we have perfect competition, then

$$\frac{w_s}{w_u} = \frac{MPL_s}{MPL_u} = \frac{G_s(L_s, L_u)}{G_u(L_s, L_u)} = g\left(\frac{L_s}{L_u}\right) \tag{11.28}$$

If we however introduce growth in L_s, then

$$Y = F(K, \gamma^t L_s, L_u) = K^\alpha G(\gamma_s^t L_s, L_u)^{1-\alpha}$$

$$\frac{w_s}{w_u} = \frac{MPL_s}{MPL_u} = \frac{\gamma_s^t G_s(L_s, L_u)}{G_u(L_s, L_u)} = g\left(\frac{L_s}{L_u}\right)\gamma_s^t \tag{11.29}$$

By introducing a shift factor γ_s^t we can model "skill-biased" technological changes. This can be empirically examined by the regression of the form:

$$\log\frac{w_s}{w_u} = \alpha + bt + c\left(\frac{L_s}{L_u}\right)$$

Suppose now that the production function has the following appearance:

$$Y = F(K, L_s, L_u) = (K + L_u)^\alpha {L_s}^{1-\alpha} \tag{11.30}$$

Then unskilled labor becomes a perfect substitute for capital and together they form a composite that is a complement of L_s.

$$\frac{w_s}{w_u} = \frac{MPL_s}{MPL_u} = \frac{(1-\alpha)(K+L_u)^\alpha {L_s}^{-\alpha}}{\alpha (K+L_u)^{\alpha-1} {L_s}^{1-\alpha}} = \frac{(1-\alpha)}{\alpha}\frac{K+L_u}{L_s} \tag{11.31}$$

If L_s goes up, the wage of skilled workers will go down. But if K goes up, then $\frac{w_s}{w_u}$ will also go up. Hence perhaps this capital-skill complementarity could help us understand what had occurred in the 1970's.

11.4.1 Changes in K

Consider an economy with two sectors:

$$c = zK_c^\alpha {L_c}^{1-\alpha} \tag{11.32}$$
$$i = qK_i^\alpha {L_i}^{1-\alpha} \tag{11.33}$$

Where z is a sector-neutral technological change and q is a investment specific technological change. Let

$$K_c + K_i = K \tag{11.34}$$
$$L_c + L_i = L \tag{11.35}$$

Combining the aggregates,

$$c + \frac{i}{q} = zK^\alpha L^{1-\alpha} \tag{11.36}$$

If we define

$$\frac{i}{q} = \frac{p_i}{p_c} \tag{11.37}$$

we have a relative price of investment goods, and this measure can be found in the data. Hence you get a measurement that captures different aspects then if you measure in the aggregates. Note that if we set $q = 1$, we have a one good economy. *[Here graphs of* $\frac{w_s}{w_u}$ *against* $\frac{L_s}{L_u}$ *for Europe and the U.S. were drawn.]*

Chapter 12

Asset pricing

The objective of this chapter is to introduce the asset pricing formula developed by Lucas (1978). We will study the pricing of assets that is consistent with the neoclassical growth model. More generally, this is the pricing methodology that is implied by the "microfoundations" approach to macroeconomics.

In the first section we go over basic Lucas' model. He works out his formula using an endowment economy inhabited by one agent. The reason for doing so is that in such an environment the allocation problem is trivial; therefore, only the prices that support a no-trade general equilibrium need to be solved for.

In the second section, we study the application of the Lucas pricing formula by Mehra and Prescott (1985). The authors utilized the tools developed by Lucas (1978) to determine the asset prices that would prevail in an economy whose endowment process mimicked the consumption pattern of the United States economy during the last century. They then compared the theoretical results with real data. Their findings were striking and are referred to as the "equity premium puzzle".

12.1 Lucas' Asset Pricing Formula

12.1.1 The Model

The representative agent in Lucas' economy solves:

$$\max_{\{c_t(z^t)\}_{t=0}^{\infty},\,\forall z^t} \left\{ \sum_t \sum_{z^t} \pi\left(z^t\right) u\left(c_t\left(z^t\right)\right) \right\}$$

$$\text{s.t.} \;\; \sum_t \sum_{z^t} p_t\left(z^t\right) c_t\left(z^t\right) = \sum_t \sum_{z^t} p_t\left(z^t\right) \omega_t\left(z^t\right)$$

$$c_t\left(z^t\right) = \omega_t\left(z^t\right) \quad \forall t, z^t \quad \text{(market clearing)}.$$

The last condition is the feasibility condition. Notice that it implies that the allocation problem is trivial, and only the prices $p_t\left(z^t\right)$ supporting this allocation as a (competitive) equilibrium must be found. (*Note*: Lucas' paper uses continuous probability.)

The asset pricing problem in Lucas' economy can be split into two parts:

1. Find an expression for $p_t(z^t)$ in terms of the primitives.
2. Apply the resulting formula $p_t(z^t)$ to price arbitrary assets.

12.1.2 Solving for prices of state-contingent claims

First-order conditions from the consumer's problem are:

$$c_t\left(z^t\right) : \beta^t \pi\left(z^t\right) u'\left(c_t\left(z^t\right)\right) = \lambda p_t\left(z^t\right), \forall t, z^t,$$

where $c_t(z^t) = \omega_t(z^t)$ will need to hold, and λ will be endogenous. We can get rid of this shadow value of income by normalizing $p_0 = 1$:

$$c_0 : u'(\omega_0) = \lambda p_0 \equiv \lambda.$$

Then

$$p_t\left(z^t\right) = \beta^t \pi\left(z^t\right) \frac{u'\left[\omega_t\left(z^t\right)\right]}{u'(\omega_0)}. \tag{12.1}$$

The Lucas Pricing Formula (12.1) shows that $p_t(z^t)$ is the price of a claim on consumption goods at t that yields 1 unit if the realized state is z^t, and 0 units otherwise.

We can distinguish three separate components in the price of this claim:

1. *Time*: $p_t(z^t)$ is decreasing in t (since $\beta < 1$).
2. *Likelihood*: $p_t(z^t)$ is increasing in the probability of occurrence of z^t.
3. *Marginal rate of substitution*: $p_t(z^t)$ is increasing in the marginal rate of substitution between goods at $(t,\, z^t)$ and $t = 0$ (don't forget that $p_t(z^t)$ is in fact a *relative* price).

For the case of a concave utility index $u(\cdot)$ (which represents risk averse behavior), the third effect will be *high* if the endowment of goods is *low* at $(t,\, z^t)$ relative to $t = 0$.

12.1.3 Pricing assets

Any asset is in essence nothing but a sum of contingent claims. Therefore, pricing an asset consists of summing up the prices of these rights to collect goods. You may already (correctly) suspect that the key is to properly identify the claims to which the asset entitles its owner. This involves specifying the time and state of nature in which these rights get activated, and the quantities of the respective contingent claims.

Consider a problem of pricing a one-period discount bond at $(t,\, z^t)$. A one-period discount bond is an asset that pays one unit at $t+1$ for every possible realization z^{t+1} such that $z^{t+1} = (z_{t+1},\, z^t)$ for $z_{t+1} \in Z$. The date-0 price of such an asset is given by

$$q_0^{rf}\left(z^t\right) = \sum_{z_{t+1} \in Z} \underbrace{p_{t+1}\left(z_{t+1},\, z^t\right)}_{\text{price of claim}} \cdot \underbrace{1}_{\text{quantity}}.$$

The date-t price is computed by

$$
\begin{aligned}
q_t^{rf}\left(z^t\right) &= \frac{q_0^{rf}\left(z^t\right)}{p_t\left(z^t\right)} \\
&= \frac{\sum\limits_{z_{t+1}\in Z} p_{t+1}\left(z_{t+1},\, z^t\right)\cdot 1}{p_t\left(z^t\right)}.
\end{aligned}
$$

Using (12.1) to replace $p_t\left(z^t\right)$ and $p_{t+1}\left(z_{t+1},\, z^t\right)$:

$$
\begin{aligned}
q_t^{rf}\left(z^t\right) &= \frac{\beta^{t+1}\sum\limits_{z_{t+1}\in Z}\pi\left(z_{t+1},\, z^t\right)\frac{u'\left(\omega_{t+1}\left(z_{t+1},\, z^t\right)\right)}{u'(\omega_0)}}{\beta^t\pi\left(z^t\right)\frac{u'\left(\omega_t\left(z^t\right)\right)}{u'(\omega_0)}} \\
&= \beta\sum_{z_{t+1}\in Z}\frac{\pi\left(z_{t+1},\, z^t\right)}{\pi\left(z^t\right)}\frac{u'\left(\omega_{t+1}\left(z_{t+1},\, z^t\right)\right)}{u'\left(\omega_t\left(z^t\right)\right)}.
\end{aligned}
$$

Notice that three components identified before now have the following characteristics:

1. *Time*: Only one period discounting must be considered between t and $t+1$.

2. *Likelihood*: $\dfrac{\pi\left(z_{t+1},\, z^t\right)}{\pi\left(z^t\right)}$ is the *conditional probability* of the state z_{t+1} occurring at $t+1$, given that z^t is the history of realizations up to t.

3. *Marginal rate of substitution*: The relevant rate is now between goods at $(t,\, z^t)$ and $(t+1,\, z^{t+1})$ for each possible z^{t+1} of the form $(z_{t+1},\, z^t)$ with $z_{t+1}\in Z$.

For more intuition, you could also think that $q_t^{rf}\left(z^t\right)$ is the price that would result if the economy, instead of starting at $t=0$, was "rescheduled" to begin at date t (with the stochastic process $\{z_t\}_{t=0}^{\infty}$ assumed to start at z^t).

Next we price a stock that pays out dividends according to the process $d_t\left(z^t\right)$ (a *tree* yielding $d_t\left(z^t\right)$ units of fruit at date-state $(t,\, z^t)$). The date-t price of this portfolio of contingent claims is given by

$$
\begin{aligned}
q_t^{tree}\left(z^t\right) &= \frac{\sum\limits_{s=t+1}^{\infty}\sum\limits_{z^s} p_s\left(z^s\right) d_s\left(z^s\right)}{p_t\left(z^t\right)} \\
&= \sum_{s=t+1}^{\infty}\sum_{z^s}\beta^{s-t}\frac{\pi\left(z^s\right)}{\pi\left(z^t\right)}\frac{u'\left(\omega_s\left(z^s\right)\right)}{u'\left(\omega_t\left(z^t\right)\right)} d_s\left(z^s\right) \\
&= E_t\left[\sum_{s=t+1}^{\infty}\beta^{s-t}\frac{u'\left(\omega_s\right)}{u'\left(\omega_t\left(z^t\right)\right)} d_s\right]. \qquad (12.2)
\end{aligned}
$$

Notice that the price includes the three components enumerated above, multiplied by the quantity of goods to which the asset entitles in each date-state. This quantity is the dividend process $d_t\left(z^t\right)$.

We can also write the price of the tree in a recursive way. In the deterministic case, this would mean that

$$p_t = \frac{p_{t+1} + d_{t+1}}{R_{t+1}},$$

where R_{t+1} is the (gross) interest rate between periods t and $t+1$. This is recursive because the formula for p_t involves p_{t+1}.

The uncertainty analogue to this expression is

$$\begin{aligned} q_t^{tree}\left(z^t\right) &= \sum_{z_{t+1}\in Z} \beta \frac{\pi\left(z_{t+1},\, z^t\right)}{\pi\left(z^t\right)} \frac{u'\left(\omega_{t+1}\left(z_{t+1},\, z^t\right)\right)}{u'\left(\omega_t\left(z^t\right)\right)} \left(d_{t+1}\left(z_{t+1},\, z^t\right) + q_{t+1}^{tree}\left(z_{t+1},\, z^t\right)\right) \\ &= \beta E_t\left[\frac{u'\left(\omega_{t+1}\right)}{u'\left(\omega_t\right)}\left(d_{t+1} + q_{t+1}^{tree}\right)\right]. \end{aligned} \tag{12.3}$$

You can check that (12.3) corresponds to (12.2) by iteratively substituting for $q_{t+1}^{tree}\left(z',\, z^t\right)$ and applying the law of iterated expectations. More importantly, notice that the price includes the usual three components. What about quantities? This expression can be interpreted as the price of a one-period tree that entitles to the dividend $d_{t+1}\left(z',\, z^t\right)$, plus the amount of consumption goods at $\left(t+1, \left(z_{t+1}, z^t\right)\right)$ needed to purchase the one-period tree again next period.

If you think about how this price fits into the endowment economy, then the amount $q_{t+1}^{tree}\left(z_{t+1},\, z^t\right)$ will have to be such that at date-state $\left(t+1,\, \left(z',\, z^t\right)\right)$ the consumer is marginally indifferent between purchasing the tree again, or using the proceeds to buy consumption goods.

More generally, let us define a random variable m_{t+1} called *the stochastic discount factor* or *pricing kernel*. Then, any random payoff X_{t+1} can be priced by

$$p_t = E_t\left[m_{t+1}X_{t+1}\right]. \tag{12.4}$$

This model is very general, and encompasses most of the asset pricing models. They differ in the particular functional form of m_{t+1}. For example, in Lucas' economy

$$m_{t+1} \equiv \beta \frac{u'\left(\omega_{t+1}\left(z_{t+1},\, z^t\right)\right)}{u'\left(\omega_t\left(z^t\right)\right)}.$$

You can check that the prices of a one-period discount bond and a stock satisfy (12.4) with m_{t+1} defined above. What is X_{t+1} there?

12.2 The Equity Premium Puzzle

The equity premium is the name of an empirical regularity observed in the United States asset markets during the last century. It consists of the difference between the returns on stocks and on government bonds. Investors who had always maintained a portfolio of shares with the same composition as Standard and Poor's SP500 index would have obtained, if patient enough, a return around 6% higher than those investing all their money in government bonds. Since shares are riskier than bonds, this fact should be explainable by the "representative agent's" dislike for risk. In the usual CIES utility function, the degree of risk aversion (but notice that also the inter-temporal elasticity of substitution!) is captured by the σ parameter.

Mehra and Prescott's exercise was intended to confront the theory with the observations. They computed statistics of the realization of (de-trended) aggregate consumption in the United States, and used those statistics to generate an endowment process in their model economy. That is, their endowment economy mimics the United States economy for a single agent.

Using parameters consistent with microeconomic behavior (drawn from microeconomics, labor, other literature, and "introspection"), they *calibrated* their model to simulate the response of a representative agent to the assumed endowment process. Their results were striking in that the model predicts an equity premium that is significantly lower than the actual one observed in the United States. This incompatibility could be interpreted as evidence against the neoclassical growth model (and related traditions) in general, or as a signal that some of the assumptions used by Mehra and Prescott (profusely utilized in the literature) need to be revised. It is a "puzzle" that the actual behavior differs so much from the predicted behavior, because we believe that the microfoundations tradition is essentially correct and should provide accurate predictions.

12.2.1 The Model

The economy is modelled as in the Lucas (1978) paper. It is an endowment economy, inhabited by a representative agent, and there are complete markets. The endowment process is characterized by two parameters that were picked so that the implied statistics matched aggregate US consumption data between 1889 and 1978.

Preferences

Preferences are modelled by the utility function

$$U = E_0\left[\sum_{t=0}^{\infty}\beta^t u\left(c_t\right)\right],$$

where the utility index u is of the CIES type:

$$u\left(c\right) = \frac{c^{1-\sigma}-1}{1-\sigma}.$$

Preferences involve two parameters, β and σ, the values of which need to be calibrated (and play an essential role in the "puzzle"). β measures the time impatience of agents. What does σ measure? In a deterministic environment, σ^{-1} is the coefficient of intertemporal substitution. But in the uncertainty case, σ is also the coefficient of relative risk aversion (CRRA):

$$CRRA \equiv -\frac{u''(c)c}{u'(c)} = -\frac{-\sigma c^{-1-\sigma}c}{c^{-\sigma}} = \sigma.$$

Therefore, the same parameter measures two (distinct) effects: the willingness to substitute consumption over time, and also across states of nature. The higher σ, the less variability the agent wants his consumption pattern to show, whatever the source of this variability: deterministic growth, or stochastic deviation.

Endowment Process

Let y_t denote income (in equilibrium, consumption) at time t. Let that y_t evolve according to

$$y_{t+1} = x_{t+1}y_t,$$

where x_{t+1} is a random variable that can take n values, $x_{t+1} \in \{\lambda_1, ..., \lambda_n\}$. The stochastic process for x_{t+1} is modelled by a first-order Markov chain, where:

$$\phi_{ij} \equiv \Pr\left[x_{t+1} = \lambda_j \left| x_t = \lambda_i\right.\right].$$

Asset Prices

Applying the Lucas' pricing formula to the *tree* that yields $d_t = y_t$ at time t, we have that

$$p_t^e = E_t\left[\sum_{s=t+1}^{\infty} \beta^{s-t}\frac{u'(y_s)}{u'(y_t)}d_s\right].$$

We will solve for these prices using a recursive formulation. First, observe that, due to the first-order Markov assumption on x_t, the likelihood of changing states is invariant over time; therefore, we can drop the time subscript and write p^e as a function of the state. Second, all information about the state of the economy at a given time can be summarized by the level of the endowment process, y_t, and the last realization of the shock, x_t. So we guess that prices will end up being a function of those two variables only. The reason why y_t is informative is that, since in equilibrium consumption is equal to endowment, y_t will provide the level of marginal utility against which future consumption streams will be compared when setting prices. x_t conveys information on the current state of the Markov process (only x_t is relevant, and not lagged values of x_t, because of the first-order assumptions). Then, the recursive formulation of the price of equity is:

$$p_e\left(x_t,\, y_t\right) = E\left[\sum_{s=t+1}^{\infty} \beta^{s-t}\left(\frac{y_t}{y_s}\right)^{\sigma} y_s \right| \left. x_t,\, y_t\right].$$

(Notice that the time subscript has been dropped from the expectation operator due to the Markov first-order assumption.)

Let us approach the solution recursively. For each state x_i, $i = 1, ..., n$, this price (at any date t) is given by:

$$p_i^e(y) = \beta \sum_{j=1}^{n} \phi_{ij} \left(\frac{y}{y\lambda_j}\right)^{\sigma} \left[y\lambda_j + p_j^e(y\lambda_j)\right] \quad \forall y, \forall i,$$

where $p_j^e(y\lambda_j)$ will be the price of equity next period if the realized state is j when consumption (endowment) growth will be $x_{t+1} = \lambda_j$.

We guess a linear solution to this functional equation:

$$p_i^e(y) = p_i^e y.$$

This yields a system of equations, with the unknowns being the coefficients p_i^e:

$$\begin{aligned} p_i^e &= \beta \sum_{j=1}^{n} \phi_{ij} (\lambda_j)^{-\sigma} \left[\lambda_j + p_j^e \lambda_j\right] \\ &= \beta \sum_{j=1}^{n} \phi_{ij} (\lambda_j)^{1-\sigma} \left[1 + p_j^e\right]. \end{aligned}$$

This equation relating p_i^e to the (weighted) summation of the p_j^e needs to hold for all i, so we have a linear system of n equations and n unknowns.

Similarly, the price of a risk-free asset paying off one unit in every state is given by

$$p_i^{rf}(y) = \beta \sum_{j=1}^{n} \phi_{ij} \lambda_j^{-\sigma} \cdot 1.$$

Notice that the level of the endowment y does not enter this formula, whereas it did enter the formula for equity prices.

Returns on Assets

Given the prices, we can compute the returns that an investor would receive by purchasing them. This will be a random variable induced by the randomness in prices and (in the case of equity) by the variability of the endowment process also. The (net) return realized at state j by an investor who purchased equity in state i is given by:

$$r_{ij}^e = \frac{\left(1 + p_j^e\right)\lambda_j}{p_i^e} - 1.$$

To understand where this formula comes from, just multiply through by y :

$$\frac{\left(1 + p_j^e\right)\lambda_j y}{p_i^e y} - 1 = \frac{\lambda_j y + p_j^e \lambda_j y - p_i^e y}{p_i^e y} \equiv \frac{d_{t+1,j} + p_{t+1,j} - p_{t,i}}{p_{t,i}}$$

The amount $d_{t+1} + p_{t+1}$ is the payoff from the tree next period (if the state is j). By subtracting the investment size $p_{t,i}$, the numerator yields the net result received by the investor. Dividing by the invested amount gives the (net) rate of return.

The conditional expected return is

$$r_i^e = E_i\left[r_{ij}^e\right] = \sum_{j=1}^n \phi_{ij} r_{ij}^e,$$

and the unconditional expected return is

$$r^e = E\left[r_i^e\right] = \sum_{i=1}^n \pi_i r_i^e = \sum_{i=1}^n \pi_i \sum_{j=1}^n \phi_{ij} r_{ij}^e.$$

r^e is not a random variable. It is an expectation taken with respect to the invariant (long run) distribution π_i of the Markov process. Recall that this is the probability vector that satisfies:

$$\Pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_n \end{pmatrix} = \Phi'\Pi.$$

In the model, p_t^e is the price of equity, that is, of the "market portfolio". It is the price of contingent claims on the whole product of the economy at t. It should be interpreted as the value of the portfolio of claims on all possible productive investments. The closest measure that we have of such a portfolio is the stock market, where shares of companies involved in almost all productive activities are traded. Therefore, r^e will be compared to actual, long run return on equity taken from the US data.

The equity premium will be given by r^e minus the long run return on government bonds (proxy for risk-free assets). In the model, (net) return on the risk-free assets is given by:

$$r_i^{rf} = \frac{1}{p_i^{rf}} - 1.$$

This is a random variable. The long run return is:

$$r^{rf} = \sum_{i=1}^n \pi_i r_i^{rf}.$$

The US data shows the following value of the equity premium:

$$r^e - r^{rf} \approx 6\%,$$

where r^e is the average return on the S&P500 from 1889 to 1978, and r^{rf} is the average yield on government bonds throughout that period.

12.2.2 Calibration

Mehra and Prescott calibrate the Markov process assuming that there are two states: $n = 2$. The values of possible endowment growth rates are:

$$\begin{aligned}\lambda_1 &= 1 + \mu + \delta \\ \lambda_2 &= 1 + \mu - \delta,\end{aligned}$$

where μ is the average growth rate $\frac{c_{t+1}-c_t}{c_t}$. Its value was chosen to be $\mu = .018$, to match that of aggregate consumption growth in the US in the period under study. δ is the variation in the growth rate.

The transition matrix was assumed symmetric, so that the probability of changing state are the same at each state:

$$\Phi = \begin{pmatrix} \phi & 1-\phi \\ 1-\phi & \phi \end{pmatrix}.$$

Then δ and ϕ are picked so as to match:

- the standard deviation of $\frac{c_{t+1}-c_t}{c_t}$, equal to .036
- the first-order serial correlation of $\frac{c_{t+1}-c_t}{c_t}$, equal to -.14.

The resulting parameter values are: $\delta = .036$, and $\phi = .43$

The remaining parameters are β and σ, which represent preferences. A priori, by introspection economists believe that β must lie in the interval (0, 1). With respect to σ, Mehra and Prescott cite several different studies and opinions on its likely values. Most micro literature suggests that σ must be approximately equal to 1 (this is the logarithmic utility case). However, some economists also believe that it could take values as high as 2 or even 4. Certainly, there seems to be consensus that σ has to be lower than 10.

Then instead of picking values for β and σ, Mehra and Prescott plotted the level of equity premium that the model would predict for different, *reasonable* combinations of values. Figure 12.1 shows approximately what they have obtained (it is a reproduction of Figure 4 of the original paper).

The model can only produce the equity premium observed in actual data at the expense of a very high risk-free interest rate, or highly *unreasonable* parameter values (such as $\beta > 1$; how do you feel about your own β?). When compared to actual data, the risk premium is too low in the model, and the risk-free rate too high. In fact, these are two puzzles.

12.3 Suggested Solutions to the Puzzle

There is one "solution" that consists of solving for parameter values that will yield the same equity premium and risk free rate as the data. You may realize that by fixing one of the preference parameters, the other can be solved for these values. An example is $\sigma \approx 15$, and $\beta \approx 1.08$. Are these values *reasonable*? What can you say from introspection? Is the total sum of instantaneous utility values bounded for these parameters?

We will briefly discuss other solutions that have been proposed in the literature:

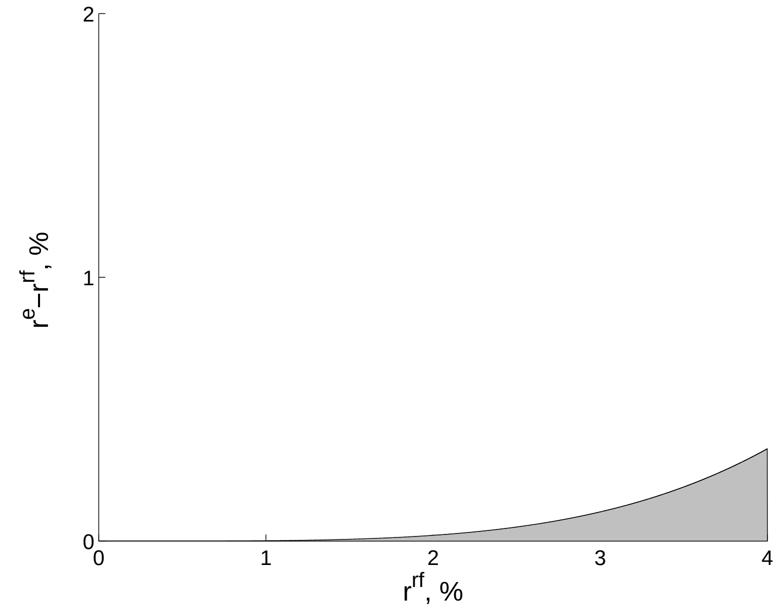

Figure 12.1: Set of admissible equity premia and risk-free returns

1. Epstein - Zin preferences. One of the issues that seem to be crucial in the puzzle is that the CIES utility function rigidly links the time structure of preferences and the aversion for risk. Both are measured by (functions of) the same parameter σ. In some sense, this is consistent with the way the risk is modelled in expected utility framework: remember that uncertainty is just the expansion of the decision making scenario to a multiplicity of "states of nature". Total utility is just the expected value of optimal decision making in each of these states. You may notice there is no difference between "time" and "states of nature". "Time" is just another subindex to identify states of the world.

   However, people seem to regard time and uncertainty as essentially different phenomena. It is natural then to seek a representation of preferences that can treat these two components of reality separately. This has been addressed by Epstein and Zin (1990), who axiomatically worked on non-expected utility and came up with the following (non-expected) utility function representation for a preference relation that considers time and states of nature as more than just two indices of the state of the world:

$$U_t = \left[c_t^{1-\rho} + \beta \left(E_t \left[U_{t+1}^{1-\sigma} \right] \right)^{\frac{1-\rho}{1-\sigma}} \right]^{\frac{1}{1-\rho}},$$

   where ρ measures inter-temporal elasticity of substitution, and σ captures risk aversion[1]. Notice that if $\rho = \sigma$, then this formula reduces to

$$U_t^{1-\rho} = c_t^{1-\rho} + \beta E_t \left[U_{t+1}^{1-\rho} \right].$$

If there is no uncertainty, then the expectation operator is redundant, and we are back to the CIES function.

This proposed solution is able to account for the risk-free rate puzzle. However, to match the equity premium it still requires an *unreasonably* high σ.

2. Habit Persistence. Suppose that each instant's utility value depends not only on current, but also on past consumption amounts (people might be reluctant to see their consumption fall from one period to the other):

$$U = E_t \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t,\, c_{t-1}\right) \right].$$

For example,

$$U = E_t \left[\sum_{t=0}^{\infty} \beta^t \frac{\left(c_t - \lambda c_{t-1}\right)^{1-\sigma}}{1-\sigma} \right].$$

This preference representation can solve the risk-free rate puzzle with *reasonable* parameter values. A related version of this type of utility function is that where utility depends on external effects (people might be happy if others around them enjoy high levels of consumption... or quite the opposite!). A possible utility index showing those characteristics could be:

$$u\left(c_t,\, \overline{c}_t,\, \overline{c}_{t-1}\right) = \frac{c_t^{1-\sigma}}{1-\sigma} \overline{c}_t^{\gamma} \overline{c}_{t-1}^{\lambda}.$$

In this example, a high value of γ can produce an equity premium value close to that in the data, with a *reasonable*, low σ. The $\overline{c}_{t-1}^{\lambda}$ component in preferences can be used to solve the risk-free puzzle. However, in spite of its ability to solve both puzzles with reasonable parameter values, this preference representation has the shortfall that it generates too variable non-stationary returns: r_i^{rf} is too variable compared to actual data, even though r^{rf} may be accurately explained.

3. Peso Problem. Suppose everybody believed that with some small probability there could be a huge disaster (a nuclear war, say). This would be accounted for in prices (and hence, returns). Such a factor might explain the equity premium.

[1]Note that it is incorrect to speak about risk aversion in the dynamic context: it measures attitude of the agent to *static* gambles. Similarly, elasticity of intertemporal substitution is not well-defined under uncertainty.

4. Incomplete Markets. A key assumption in the Mehra and Prescott model is that there is a representative agent whose consumption equals aggregate consumption. This can be generalized to a numerous population if we assume that all individuals are perfectly insured - the maximum variability their consumption can show is aggregate variability. However, it is not true that every person's consumption has exactly the same variability as aggregate consumption. Individuals' productivity could also be subject shocks by itself (for instance, becoming handicapped after an accident).

   Such a mismatch would imply that trying to explain the puzzles by a model based on a representative agent could not be successful. If markets are incomplete, equity holding decisions are taken by individuals who suffer "idiosyncratic" stochastic shocks that may differ from one another, and due to the incompleteness, consumers are not able to insure themselves against this idiosyncratic risk. Return differentials between risky and risk-free assets then must lure into equity individuals whose consumption variability is larger than the aggregate.

5. Transaction Costs. Some authors have tried to explain the high risk premium as the consequence of high transaction costs to buy shares. However, this needs unrealistic cost levels to match the data.

6. Production. Mehra and Prescott's model is an endowment economy. Could the introduction of production into the model affect its results? The answer is no: it is consumption that we are interested in; it does not really matter how consumption is provided for. This approach is not really relevant.

7. Leverage. In Mehra and Prescott's model, equity is the price of the "tree" that yields the whole economy's production. However, actual equity does not exactly give its owner rights to the whole product of a company. Other parties have rights over a company's economic surplus, that come before shareholders. Creditors' claims have priority in case of bankruptcy.

   Therefore, actual share dividends are more risky than consumption. There are "legal" risks involved in investing in shares, that are not reflected in Mehra and Prescott's formulation. Financial profession tends to believe in this explanation more than economists.

12.4 References

Epstein, Larry and Stanley Zin, "First-order risk aversion and the equity premium puzzle," *Journal of Monetary Economics*, 26(1990), 387-407.

Kocherlakota, Narayana, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34(1996), 42-71.

Lucas, Robert E. Jr., "Asset Prices in an Exchange Economy," *Econometrica*, 46(1978), 1429-45.

Mehra, Rajnish and Edward C. Prescott, "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(1985), 145-61.

Chapter 13

Economic policy

In this chapter we will study the effects of financing a given stream of government consumption. We will focus on fiscal policy, and in particular on the aspects related to funding an arbitrary sequence of government expenditure.

The government's expenditure plan will be treated as a given, and the analysis will focus on financing such a plan under a general equilibrium framework. Given a sequence $\{g_t\}_{t=0}^{\infty}$ of government expenditures, what can our microfoundations approach to macroeconomics tell us about the best way to provide funding for it?

In particular, we will examine the following two questions:

1. *Q*: If it is "technologically" possible to implement lump-sum taxation, does the timing of these taxes matter? If so, how?

   A: The *Ricardian equivalence* tells us that timing of lump-sum taxes does not matter. This holds unconditionally in the dynastic model, but in the overlapping-generations setup it depends on preferences.

2. *Q*: If lump-sum taxes are not enforceable, what kinds of distortionary taxes are the best? What can we say about timing of taxation in this case?

   A: The answer to this issue is much less straightforward than the previous one. The most widely mentioned distortionary taxes in the literature are levied on factor remunerations. We will analyze the case of proportional taxes on labor income (τ_t), and on capital income (θ_t).

   A sequence of proportional taxes $\{\tau_t, \theta_t\}_{t=0}^{\infty}$ has to be chosen so as to optimize some measure of welfare (i.e., to pick the best allocation for some given ranking of outcomes). But, besides the issue of how to properly assess welfare, an important issue arising is that of the "time-consistency" of a proposed taxing sequence (the literature on this topic, as so many others, was originated by Prescott).

   Usually, models predict that the best distortionary taxing policy is to fully tax initial capital. Intuitively, since capital that has been invested is a "sunk" cost and cannot escape the taxing, this is the least distortionary tax, provided that the government could *credibly* commit to implement this tax *only once*. However, in the usual stationary model, at $t = 1$ the government's problem is identical to that at $t = 0$ (only the initial capital and maybe the history of shocks will differ). Hence, whatever was optimal at $t = 0$ will again be optimal at $t = 1$. So a promise on the part of the government fully tax capital at $t = 0$ only and never again could not be rationally believed - we say that it would not be time-consistent.

13.1 Ricardian equivalence

We will analyze the welfare effects of timing in lump-sum taxation by adding the presence of a government to our usual macro model. This presence takes the form of a given expenditure sequence $\{g_t\}_{t=0}^{\infty}$ and an ability to levy taxes on the consumer's income. We begin with the dynastic model, and then analyze the overlapping-generations setup.

In both cases, what we are interested in is finding the sequence of debt $\{B_t\}_{t=0}^{\infty}$ (one-period loans from the private sector to the government) and lump-sum taxes $\{\tau_t\}_{t=0}^{\infty}$ such that the following budget constraint is satisfied at every t:

$$g_t + B_{t-1} = q_t B_t + \tau_t, \forall t. \tag{13.1}$$

Equation (13.1) requires that sources and uses of funds be equalized in every period. Funds are used to finance expenditures g_t, and to repay B_{t-1} (bonds issued at $t-1$ that must be redeemed at t). Sources are lump-sum tax collection τ_t and new government borrowing B_t. q_t is the price of these bonds - the amount of "money" (in this case, notice that the numeraire is g_t, which will turn out to be consumption goods) that the government gets for each unit of bonds B_t issued. This price is just the inverse of the (gross) return on these bonds. We will assume that the initial bond position is null: $B_{-1} = 0$.

13.1.1 Dynastic model

Preferences will be assumed strongly monotone so that consumers will exhaust their budget constraints at every period. Consumption goods are provided by an exogenous, deterministic endowment process. The problem will be formulated sequentially; since there is one state of the world for each t, just one asset per period is enough for complete markets to obtain. In addition to government bonds, agents will be allowed to hold positions in one-period loans; i.e., they will be able to borrow or lend for one period at each t.

We write down the consumer's sequential budget constraint in the dynastic economy:

$$c_t + q_t B_t + l_t = \omega_t + B_{t-1} + l_{t-1} R_t - \tau_t,$$

where l_t denotes the net borrowing/lending position at the end of period t; ω_t is the exogenously given endowment of consumption goods at t.

We assume that a no-Ponzi-game condition holds. Then, using equivalence of sequential and date-0 formulations, we may consolidate the budget constraint to

$$\sum_{t=0}^{\infty} p_t c_t = \sum_{t=0}^{\infty} p_t \omega_t - \sum_{t=0}^{\infty} p_t \tau_t + \sum_{t=0}^{\infty} \left(p_{t+1} - p_t q_t\right) B_t.$$

p_t is date-0 price of a unit of consumption good at t. We can normalize $p_0 = 1$, and we also have that

$$\frac{p_t}{p_{t+1}} \equiv R_{t+1}.$$

In equilibrium, government and private debt must yield the same rate of return, or otherwise asset markets would not clear:

$$q_t = \frac{1}{R_{t+1}}.$$

This implies that

$$p_{t+1} - p_t q_t = 0.$$

There is only one state of the world (this is a deterministic model), but there are two assets. So, one of them must be redundant.

Replacing in the consumer's budget constraint, we obtain

$$\sum_{t=0}^{\infty} p_t c_t = \sum_{t=0}^{\infty} p_t \omega_t + \sum_{t=0}^{\infty} p_t \tau_t.$$

The government's consolidated budget reads

$$\sum_{t=0}^{\infty} p_t g_t = \sum_{t=0}^{\infty} p_t \tau_t.$$

But then if we substituted this in the consumer's budget constraint, we would realize that in fact what is relevant for the decision making agent is not the taxing stream, but the expenditure stream $\{g_t\}_{t=0}^{\infty}$:

$$\sum_{t=0}^{\infty} p_t c_t = \sum_{t=0}^{\infty} p_t \omega_t + \sum_{t=0}^{\infty} p_t g_t. \tag{13.2}$$

Equation (13.2) is the "Ricardian Equivalence": the timing of taxes is not relevant. For a more formal statement of this equivalence, we first need to define the competitive equilibrium in this economy:

Definition 13.1 *A* ***competitive equilibrium*** *is a sequence* $\{p_t,\, c_t,\, (g_t)\,,\, B_t,\, q_t,\, r_t,\, \tau_t\}_{t=0}^{\infty}$ *such that:*

1. *Consumers' utility is maximized, subject to their budget constraint.*

2. The government's budget constraint is satisfied.

3. Markets clear. In the case of an endowment economy, this condition requires that $c_t + g_t = \omega_t$. In the case of a production economy, it requires that $c_t + K_{t+1} + g_t = F(K_t, n_t)$[1].

(4. Firms maximize profits - in the case of a production economy.)

Notice that in naming the sequence $\{p_t, c_t, (g_t), B_t, q_t, r_t, \tau_t\}_{t=0}^{\infty}$, we have written the government's expenditure stream g_t in parentheses. The reason is that in fact this is given, and as such is not a decision variable that should be part of the equilibrium. It could be treated as a parameter in the problem (for example, in an endowment economy the endowment could be redefined as net of government expenditures).

Notwithstanding the way government expenditures are presented in the definition, equipped with a competitive equilibrium we are now ready to state the following:

Theorem 13.2 (Ricardian equivalence in a dynastic model) *Let the sequence $\{p_t, c_t, g_t, B_t, q_t, r_t, \tau_t\}_{t=0}^{\infty}$ be an equilibrium. Then $\left\{p_t, c_t, g_t, \widehat{B}_t, q_t, r_t, \widehat{\tau}_t\right\}_{t=0}^{\infty}$ is also an equilibrium if*

$$\sum_{t=0}^{\infty} p_t \widehat{\tau}_t = \sum_{t=0}^{\infty} p_t \tau_t$$

and the sequence $\left\{\widehat{B}_t\right\}_{t=0}^{\infty}$ is picked to satisfy the government's budget constraint:

$$\widehat{B}_t q_t - \widehat{B}_{t-1} + \widehat{\tau}_t = B_t q_t - B_{t-1} + \tau_t.$$

Proof. The proof of this theorem is immediate from equation (13.2). The new tax-borrowing mix chosen by the government does not alter the actual budget constraint faced by the consumer. And since his maximization problem remains completely unaltered, the optimizing choices remain the same. ■

13.1.2 Overlapping generations

The overlapping-generations setup seems more suitable to highlight the inter-generational transfer aspects that may be involved when government picks a tax-borrowing mix to finance expenditures. Changes in timing will alter the budget constraints faced by consumers if different generations are involved in the change. The actual effect of this on present and future generations' well-being will depend on the extent to which the current generation values the welfare of its offspring.

[1] $F(K_t, n_t)$ is assumed to incorporate depreciation of capital.

Hence, a limited version of the Ricardian equivalence result holds in this case. In order to analyze it, we will have to make use of a competitive equilibrium. In this case, it will take the form of a sequence

$$\{c_t(t),\, c_t(t+1),\, g_t,\, l_t,\, B_t,\, \tau_t(t),\, \tau_t(t+1),\, r_t,\, q_t\}_{t=0}^{\infty},$$

such that the conditions for it to be an equilibrium are satisfied. Then we can state the following:

Theorem 13.3 (Ricardian equivalence in an overlapping-generations model) *Let the sequence* $\{c_t(t),\, c_t(t+1),\, g_t,\, l_t,\, B_t,\, \tau_t(t),\, \tau_t(t+1),\, r_t,\, q_t\}_{t=0}^{\infty}$ *be an equilibrium. Then so is* $\left\{c_t(t),\, c_t(t+1),\, g_t,\, l_t,\, \widehat{B}_t,\, \widehat{\tau}_t(t),\, \widehat{\tau}_t(t+1),\, r_t,\, q_t\right\}_{t=0}^{\infty}$*, where*

$$\tau_t'(t) + \frac{\tau_t'(t+1)}{r_{t+1}} = \tau_t(t) + \frac{\tau_t(t+1)}{r_{t+1}}, \forall t$$

and

$$q_t B_t' - B_{t-1}' + \tau_t'(t) + \tau_{t-1}'(t) = q_t B_t - B_{t-1} + \tau_t(t) + \tau_{t-1}(t) = g_t, \forall t.$$

Proof. You may notice that the theorem states the equivalence for the case where the present value of taxation is not changed for any generation. The argument, therefore, will be of the same nature as the dynastic case. First recall that the consumer's sequential budget constraints are

$$c_t(t) + q_t B_t + l_t = \omega_t(t) - \tau_t(t)$$
$$c_t(t+1) = B_t + r_{t+1} l_t + \omega_t(t+1) - \tau_t(t+1).$$

Substituting the suggested alternative tax-borrowing scheme, it is easy to see that each generation's budget set remains unaltered. Then so are the agents' choices. Figure 13.1 illustrates that the change in taxes implies no alteration to the actual budget constraint. The point c^* will be chosen, regardless of whether the taxes are $(\tau_t(t), \tau_t(t+1))$ or $(\tau_t'(t), \tau_t'(t+1))$. The slope of the budget line is $-r_{t+1}$. ■

Next, let us suppose that the old at t care about the utility enjoyed by the young at t. And let government collect a given amount of taxes, choosing between taxing young at 0 or old at 0 (we could equally well take any generations t and $t+1$). Will the choice of taxation have an impact in total welfare?

We assumed that the utility of the old at 0 is a function of the utility of their offspring:

$$u_{-1}\left[c_{-1}(0),\, u_0\left(c_0(0),\, c_0(1)\right)\right].$$

Government's budget constraint requires that:

$$\tau_{-1}(0) + \tau_0(0) = g_0.$$

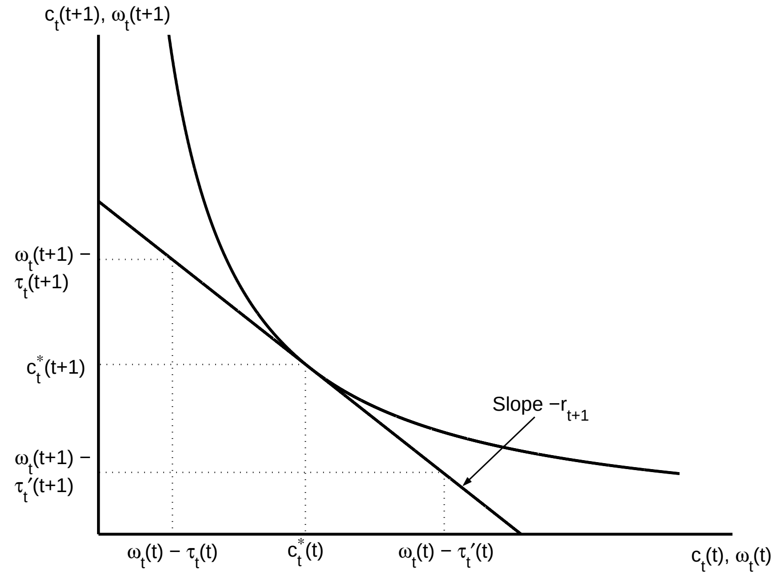

Figure 13.1: Generation t's budget constraint under different tax-borrowing schemes

The private budgets for generations -1 and 0 are:

$$\begin{aligned} c_{-1}(0) &= \omega_{-1}(0) - \tau_{-1}(0) - b_{-1} \\ l_0 + c_0(0) &= \omega_0(0) - \tau_0(0) + b_{-1} \\ c_0(1) &= r_1 l_0 + \omega_0(1), \end{aligned}$$

where $b_{-1} \geq 0$ is a *bequest* that the old leave to their descendants, and r_1 is the return on savings between periods $t = 0$ and $t = 1$.

The old at $t = 0$ solve:

$$\max_{b_{-1}} \; u_{-1}\left[\omega_{-1}(0) - \tau_{-1}(0) - b_{-1},\, u_0\left(\omega_0(0) - \tau_0(0) + b_{-1},\, \omega_0(1)\right)\right]$$
$$\text{s.t.} \quad b_{-1} \geq 0.$$

(We have used the fact that $l_t = 0$ must prevail in equilibrium in an endowment economy.)

Figure 13.2 shows the trade-off faced by the government. The slope of the straight line is -1, reflecting that every unit of extra consumption given to the young must be subtracted from the old. The point $c_{-1}^*(0)$ is the optimizing choice of the old; it implies a certain bequest $b_{-1} \geq 0$. The government can only induce a consumption choice with $b_{-1} \geq 0$; therefore, all points to the right of $\omega_{-1}(0)$ are not feasible. If the government chooses any taxation between $\tau_{-1}(0) = 0$ and $\tau_{-1}(0) = \omega_{-1}(0) - c_{-1}^*(0)$, then in fact nothing changes and the old "correct" the "bad" choice of the government through the appropriate choice of b_{-1}. However, if the government chooses a taxation mix to the left of $c_{-1}^*(0)$, then the solution to the bequest level becomes corner.

Summarizing, changes in taxation timing will yield changes in the consumption allocation (and, as a result, in welfare) whenever bequest constraints bind. Otherwise, they will not.

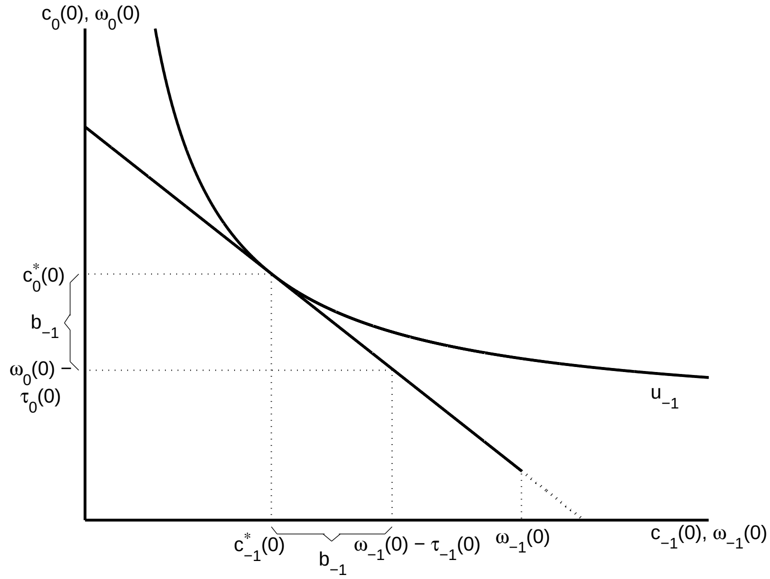

Figure 13.2: Inter-generational trade-off

13.2 Optimal distortionary taxes

Next we address the second issue raised in the introduction of this chapter - the issue of optimal distortionary taxation. We will use the standard neoclassical growth model in its dynastic version, and with endogenous labor supply (valued leisure). We will take as given a (constant) sequence of government expenditure $\{g_t\}_{t=0}^{\infty}$, with $g_t = g, \forall t$. Lump-sum taxation will not be available, and the government will levy taxes on factor remunerations: capital and/or labor income. The government will choose the optimal mix so as to maximize the representative agent's utility.

13.2.1 Taxation smoothing

A question we would like to answer in this section is what a time pattern of optimal taxes on labor would look like. We will start with a standard infinite-horizon maximization problem for consumer:

$$\max_{\{c_t, l_t, b_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u\left(c_t, 1 - l_t\right)$$
$$\text{s.t. } c_t + q_t b_{t+1} = w_t l_t \left(1 - \tau_t\right) + b_t, \forall t,$$

where b_t is government bond holdings at time t, q_t is the price of a one-period discount bond at time t, c_t and l_t are consumption and labor effort at time t respectively, and τ_t is the tax on labor income. To simplify the problem, let us assume that production uses only labor, and production function is linear in labor:

$$F\left(l\right) = l.$$

Then, in a competitive equilibrium the wage rate will be $w_t = 1, \forall t$. Assuming that $\{g_t\}_{t=0}^{\infty}$ is a sequence of public spending, government budget constraint is

$$q_t b_{t+1} + \tau_t l_t = g_t + b_t, \forall t. \tag{13.3}$$

Notice that we have built into government's budget constraint two equilibrium assumptions: first, the equilibrium wage rate $w_t = 1$, and second, the market clearing condition for bond market (what households are willing to lend is what government is willing to borrow at the going interest rate).

Suppose that private sector's problem has "sufficient strict concavity" so that the solution to the private sector problem can be characterized by first-order conditions. The first-order conditions of the private sector are:

$$u_c\left(l_t - g_t, 1 - l_t\right)\left(1 - \tau_t\right) = u_{1-l}\left(l_t - g_t, 1 - l_t\right) \tag{13.4}$$

$$q_t u_c\left(l_t - g_t, 1 - l_t\right) = \beta u_c\left(l_{t+1} - g_{t+1}, 1 - l_{t+1}\right), \tag{13.5}$$

where equation (13.4) represents consumption-leisure tradeoff, and equation (13.5) represents consumption-savings tradeoff. Notice that both equations incorporate consumption goods market clearing condition: $c_t + g_t = F\left(l_t\right) \equiv l_t$.

Now, we can combine (13.3), (13.4), and (13.5) into one expression that does not involve τ_t or q_t at all[2]:

$$\eta\left(l_t, b_t, g_t, l_{t+1}, b_{t+1}, g_{t+1}\right) \equiv u_c\left(l_t - g_t, 1 - l_t\right)\left[1 - \frac{g_t + b_t - \beta b_{t+1} \frac{u_c\left(l_{t+1} - g_{t+1}, 1 - l_{t+1}\right)}{u_c\left(l_t - g_t, 1 - l_t\right)}}{l_t}\right] - $$
$$-u_{1-l}\left(l_t - g_t, 1 - l_t\right) = 0.$$

$\eta\left(l_t, b_t, g_t, l_{t+1}, b_{t+1}, g_{t+1}\right)$ is a quite complicated and generally nonlinear function. The advantage of this formulation is that the government's problem can now be written as:

$$\begin{array}{rl} \max\limits_{\{c_t, l_t, b_{t+1}\}_{t=0}^{\infty}} & \sum_{t=0}^{\infty} \beta^t u\left(c_t, 1 - l_t\right) \\ \text{s.t.} & \eta\left(l_t, b_t, g_t, l_{t+1}, b_{t+1}, g_{t+1}\right) = 0, \forall t \\ & b_0 \text{ given.} \end{array} \tag{13.6}$$

Two issues are relevant to note regarding problem (13.6):

- Will the first-order conditions to the government problem be sufficient? In general no, even under the usual concavity assumptions for utility function. The reason is that $\eta\left(\cdot\right)$ is highly non-linear, which can potentially lead to multiple solutions and local maxima.

- The Laffer curve argument seems to be relevant here: given a government's choice of labor supply l_t there can be several ways to implement it (i.e. several τ_t that will result in the same labor supply decision, but different levels of consumption)[3].

[2]This was obtained by factoring τ_t from (13.3), substituting for q_t from (13.5), and plugging the result into (13.4) (check).

[3]This issue can be at least partially mitigated by formulating problem in terms of b_t and τ_t.

Let us ignore these issues for the moment, and suppose that first-order conditions to (13.6) characterize the solution to government problem. Consider two cases:

1. Quasilinear utility function. Suppose that period utility is given by

$$u(c, 1-l) = c + v(1-l).$$

   Then, from (13.5) we know that in equilibrium $q_t = \beta, \forall t$. We can now guess that a *constant* sequence of tax rates $\{\tau_t\}_{t=0}^{\infty} = \{\bar{\tau}\}_{t=0}^{\infty}$ satisfies first-order conditions. Indeed, given that $\tau_t = \bar{\tau}$, from (13.4) we can infer that $l_t = \bar{l}, \forall t$. Then, we can use (13.4) and government's present value budget constraint

$$\begin{aligned}\sum_{t=0}^{\infty} \beta^t g_t + b_0 &= \sum_{t=0}^{\infty} \beta^t \bar{l}\bar{\tau} \\ &= \frac{1}{1-\beta}\bar{l}\bar{\tau}\end{aligned}$$

   to solve for $\bar{\tau}$ and $\bar{l}$. This is a *tax smoothing* result: regardless of the pattern of government spending $\{g_t\}_{t=0}^{\infty}$, government should borrow and lend to preserve constant tax rate over time.

2. General utility function. With a general $u(c, 1-l)$ distortion smoothing need not follow. Moreover, in many cases we will get *time inconsistency* of government plan. To see why this will be the case, suppose that $\frac{d\eta(\cdot)}{dl_{t+1}} \neq 0$. Now consider the first-order conditions to government problem with respect to l_t. For $t > 0$ the first-order conditions will include three terms: (1) direct utility effect $u_{1-l}(c_t, 1-l_t)$; (2) constraint at t $\frac{d}{dl_t}\eta(l_t, g_t, b_t, l_{t+1}, g_{t+1}, b_{t+1})$; and (3) constraint at $t-1$ $\frac{d}{dl_t}\eta(l_{t-1}, g_{t-1}, b_{t-1}, l_t, g_t, b_t)$. In contrast, the first-order condition for $t = 0$ will include only the first two elements. Therefore, the government policy at period 0 will generally differ from the policy at $t > 0$, even if the public expenditures are constant over time. In other words, suppose that the solution to the government problem is some $\{l_t, b_{t+1}\}_{t=0}^{\infty}$. Now suppose that at period, say, $t = 100$ the government is allowed to re-optimize. Would it adhere to the plan chosen earlier? The answer is no. The reason is that by allowing the government to re-optimize we make period $t = 100$ the beginning of time, and the first-order condition to the government problem will be different from the one that yielded the original policy sequence.

   As a result, this problem cannot be formulated in dynamic programming language, since dynamic programming by construction assumes that the problem is *time-consistent*, i.e. the agent is unwilling to change their choices at a later date.

   Another perspective on time-inconsistency of the government problem is the issue of *commitment*. The solution to problem (13.6) is only valid if the government can commit to the plan described by such solution. Otherwise, given the rational expectations assumption, the consumers will foresee government's desire to deviate, and will change their actions accordingly. The problem then changes into an infinitely repeated game between the consumers and the government, and becomes substantially more complicated.

13.2.2 Optimal taxation mix

Now we will change assumptions on the production function of the economy to include capital, and allow government to levy taxes on both capital and labor. The question we will be interested in is what the optimal taxation scheme would be, i.e. at what rates capital and labor will be taxed.

This is a high-dimensional problem. The decision making agent chooses a consumption-leisure-capital accumulation path given the tax rates. So the government, when choosing taxation, has to take into account how these rates affect consumer's choices at different points in time.

We will assume that the individual's preferences over consumption and labor streams are represented by a time separable utility function with a discount factor $\beta \in (0, 1)$. This function is strictly increasing in c_t, strictly decreasing in n_t, and concave. A central planner seeking an optimal allocation in a deterministic economy would thus solve:

$$\max_{\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t u\left(c_t,\, n_t\right) \right\}$$
$$\text{s.t. } c_t + g + K_{t+1} = F\left(K_t,\, n_t\right) + \left(1-\delta\right) K_t, \forall t.$$

We want to study the outcome from distortionary taxes, so we need to decentralize the problem. Let τ_t denote the proportional tax on labor income, and θ_t that on capital income. Then we have the following:

Definition 13.4 *A* ***competitive equilibrium*** *is a sequence*

$$\{c_t,\, K_{t+1},\, n_t,\, p_t,\, r_t,\, w_t,\, \theta_t,\, \tau_t\}_{t=0}^{\infty}$$

such that:

1. *The consumption-leisure-capital accumulation path* $\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}$ *maximizes consumers' utility subject to the budget constraint*

$$\sum_{t=0}^{\infty} p_t\left(c_t + K_{t+1}\right) = \sum_{t=0}^{\infty} p_t \left[\left(1-\tau_t\right) w_t n_t + R_t^K K_t\right],$$

where $R_t^K = 1 + \left(1-\theta_t\right)\left(r_t - \delta\right)$ *denotes the gross return on capital after taxes. Notice that depreciated capital is not taxed. You can think that if* r_t *is the revenue from lending capital to producers, then* δ *is the cost that the owner of capital faces due to depreciation. The tax is only levied on the net income from capital.*

2. Firms maximize profits:

$$\{K_t,\, n_t\}_{t=0}^{\infty} = \underset{\left\{\tilde{K}_t, \tilde{n}_t\right\}_{t=0}^{\infty}}{\arg\max} \left\{F\left(\tilde{K}_t, \tilde{n}_t\right) - w_t \tilde{n}_t - r_t \tilde{K}_t\right\}.$$

3. The government's budget constraint is satisfied:

$$\sum_{t=0}^{\infty} p_t g = \sum_{t=0}^{\infty} p_t \left[\tau_t w_t n_t + \theta_t \left(r_t - \delta\right) K_t\right].$$

4. Markets clear:

$$c_t + g + K_{t+1} = F\left(K_t,\, n_t\right) + \left(1 - \delta\right) K_t, \forall t.$$

We will first focus on studying this problem in the steady state, i.e. when $\theta_t = \theta$, $\tau_t = \tau$, $c_t = c$, $n_t = n$, and $K_t = K$ for all t. Consider the consumer's problem. The first order conditions are:

$$\begin{aligned}
c_t &: \ \beta^t u_c\left(c_t,\, n_t\right) = \lambda p_t \\
K_{t+1} &: \ p_t = R_{t+1}^K p_{t+1} \\
n_t &: \ -\beta^t u_n\left(c_t, n_t\right) = \left(1 - \tau_t\right) w_t \lambda p_t.
\end{aligned}$$

Rearranging,

$$\begin{aligned}
u_c\left(c_t,\, n_t\right) &= \beta u_c\left(c_{t+1},\, n_{t+1}\right) R_{t+1}^K \\
u_n\left(c_t, n_t\right) &= -\left(1 - \tau_t\right) w_t u_c\left(c_t, n_t\right).
\end{aligned}$$

Given steady state assumption, we have $R_t^K = R^K$, and it must satisfy

$$\beta R^K = 1$$

and

$$R^K = 1 + \left(1 - \theta\right)\left(r - \delta\right),$$

where r is the steady state factor payment. Assuming that production technology exhibits constant returns to scale, $r = F_1\left(K,\, n\right)$ is consistent with equilibrium and with market clearing. In addition, under this assumption F_1 is a function of $\frac{K}{n}$. Therefore,

$$R^K = 1 + \left(1 - \theta\right)\left[f_1\left(\frac{K}{n}\right) - \delta\right],$$

and we can solve for $\frac{K}{n}$ (notice that this ratio will depend on the tax policy θ). To solve for labor, we will use the first order conditions with respect to that decision variable, and those involve the corresponding tax rate τ.

Next we turn to the government. Its decision problem amounts to choosing the sequence

$$\pi = \{\pi_t\}_{t=0}^{\infty} \equiv \{\theta_t,\, \tau_t\}_{t=0}^{\infty}$$

in order to maximize the consumer's welfare, while satisfying the government's budget constraint. The solution to the individuals' problem showed that the tax choice will induce an optimal behavior on the part of the consumer as a function of that choice. Therefore, we may define for every tax sequence π an allocation rule

$$x(\pi) = \{c_t,\, K_t,\, n_t\}_{t=0}^{\infty}$$

that comes from the consumer's response to the tax sequence π.

The taxing policy also determines the sequence of prices:

$$w(\pi) = \{p_t,\, r_t,\, w_t\}_{t=0}^{\infty}.$$

These are the prices supporting $x(\pi)$ as a competitive equilibrium.

Then for any π, there is a competitive equilibrium which from this point onwards we will denote by

$$\begin{aligned} x(\pi) &= \{c_t(\pi),\, K_t(\pi),\, n_t(\pi)\}_{t=0}^{\infty} \\ w(\pi) &= \{p_t(\pi),\, r_t(\pi),\, w_t(\pi)\}_{t=0}^{\infty}. \end{aligned}$$

With these elements, we can introduce a useful tools to study this problem:

Definition 13.5 (Ramsey equilibrium) *A* ***Ramsey equilibrium*** *is a tax policy π (that the government chooses optimally so as to be in budgetary equilibrium), an allocation rule $x(\pi)$, and a price rule $w(\pi)$ such that:*

(i) π maximizes:

$$\sum_{t=0}^{\infty} \beta^t u\left(c_t(\pi),\, n_t(\pi)\right)$$

subject to the government's budget constraint and with allocations and prices given by $x(\pi)$ and $w(\pi)$.

(ii) For every alternative policy π', $x(\pi')$ and $w(\pi')$ constitute a competitive equilibrium given policy π'.

(iii) $\theta_0 = \overline{\theta}_0$.

This is an important restriction. The initial level of tax on capital income must be exogenously given. Otherwise, if the government could choose θ_0 arbitrarily high, and $\tau_t = \theta_t = 0\ \forall t \geq 1$, taxing initial capital would be like a lump-sum tax, since initial capital is essentially a "sunk" investment, which cannot be modified.

There are two approaches to this problem.

(I) The government directly solves

$$\max_{\{c_t,\, n_t,\, K_{t+1},\, \theta_t,\, \tau_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t u\left(c_t,\, n_t\right) \right\}$$

$$\text{s.t.} \quad \left[\begin{array}{l} \beta^t u_c\left(c_t,\, n_t\right) = \lambda p_t \\ \beta^t u_n\left(c_t,\, n_t\right) = -\lambda p_t\left(1-\tau_t\right) w_t \\ p_t = R_{t+1}^K p_{t+1} = \left[1+\left(1-\theta_{t+1}\right)\left(r_{t+1}-\delta\right)\right] p_{t+1} \\ r_t = F_K\left(K_t,\, n_t\right) \\ w_t = F_n\left(K_t,\, n_t\right) \end{array} \right. \tag{13.7}$$

$$c_t + g + K_{t+1} = F\left(K_t,\, n_t\right) + \left(1-\delta\right) K_t \tag{13.8}$$

$$\sum_{t=0}^{\infty} p_t g = \sum_{t=0}^{\infty} p_t\left[\tau_t w_t n_t + \theta_t\left(r_t-\delta\right) K_t\right], \tag{13.9}$$

where (13.7) are the first order conditions which, together with the market clearing conditions (13.8), define a competitive equilibrium; and (13.9) is the government's own budget constraint.

(II) Instead of the previous huge system, we could solve the problem in a smarter way by having government solve:

$$\max_{\{c_t, K_{t+1}, n_t\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t u\left(c_t,\, n_t\right) \right\}$$

$$\begin{array}{ll} \text{s.t.} & c_t + g + K_{t+1} = F\left(K_t,\, n_t\right) + \left(1-\delta\right) K_t, \forall t \\ & \sum\limits_{t=0}^{\infty} \beta^t\left[u_c\left(c_t,\, n_t\right) c_t + u_n\left(c_t,\, n_t\right) n_t\right] = u_c\left(c_0,\, n_0\right) R_0^K K_0, \end{array}$$

where $R_0^K = 1 + \left[1-\overline{\theta}_0\right]\left[F_K\left(K_0,\, n_0\right)-\delta\right]$. We will call the second constraint in this formulation the "implementability" constraint.

The claim is that solving the problem (II) is equivalent to solving (I). Then the two constraints in (II) must contain the same information as the huge system of constraints in (I).

In addition, notice that in the system (II), the government's decision variables are not the tax sequence π anymore, but directly the consumption-capital accumulation-labor supply path $\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}$. Thus, for the two problems to be equivalent, it must be the case that by choosing these three paths subject to the two constraints in (II), we must be indirectly choosing all other variables, in particular taxes.

This means that any sequence $\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}$ satisfying the two constraints in (II) has to be a part of a competitive equilibrium vector. We will now show that this is true. Define prices using the usual guesses:

$$
\begin{aligned}
r_t &= F_K\left(K_t,\, n_t\right) \\
w_t &= F_n\left(K_t,\, n_t\right) \\
p_0 &= 1 \\
p_t &= \beta^t \frac{u_c\left(c_t,\, n_t\right)}{u_c\left(c_0,\, n_0\right)}.
\end{aligned}
$$

Let the taxes on labor income be determined by the equation

$$
\left(1-\tau_t\right) F_n\left(K_t,\, n_t\right) = -\frac{u_n\left(c_t,\, n_t\right)}{u_c\left(c_t,\, n_t\right)},
$$

and the taxes on capital income by

$$
u_c\left(c_t,\, n_t\right) = \beta u_c\left(c_{t+1},\, n_{t+1}\right)\left[1+\left(1-\theta_{t+1}\right)\left(F_K\left(K_{t+1},\, n_{t+1}\right)-\delta\right)\right]. \tag{13.10}
$$

So, are the conditions for a competitive equilibrium met?

- *Market clearing*: Yes, since $\{c_t,\, K_{t+1},\, n_t\}_{t=0}^{\infty}$ was assumed to satisfy the two restrictions in (II), and one of those was precisely market clearing.
- *Consumers' and firms' first order conditions*: Yes, they are satisfied. This follows from our guesses for prices and taxes (check).
- *Individuals' budget constraint*: If we use the second restriction in (II) and substitute prices and taxes back in, then this restriction will become exactly an individual's budget constraint (check).
- *Government's budget constraint*: If individual's budget constraints are met, and markets clear, then we must have that the government's constraint is also met. This argument is similar to a Walras' law type of reasoning.

It looks like we have immensely simplified system (I) into system (II). However, this is not for free. Two drawbacks from the alternative approach to the problem must be highlighted:

1. The constraint set looks "weirder" than in our usual maximization problem. In particular, the equation in the second constraint might have a very arbitrary shape. The requirements for sufficiency of the first order conditions, therefore, will not necessarily be met. Points solving problem (II) will have to be cross-checked to make sure that they maximize the objective function.
2. Do you think that it is possible to apply dynamic programming techniques to solve (II)? Is it possible to write this as a recursive problem? Unfortunately, the answer is <u>no</u>. It is not possible to formulate (II) recursively.

Notice that second drawback that we have mentioned goes beyond the mathematical aspects involved. What does the impossibility of formulating (II) recursively tell us about the economic nature of the problem we are dealing with? The answer is that the problem is not stationary, because any solution to it cannot be time-consistent. If we rename any $t > 0$ as $t = 0$, and act as if the economy was starting at that point, then the government would be willing to *revise* the decisions taken originally for that t, in particular, the decision regarding taxing capital income.

This implies that any solution to (II) is a non-credible promise on the part of the government, since it will be willing to modify its original plan at every point in time. The way we overcome this drawback is that we assume that there is some sort of *commitment device* (enforceable laws, for example), which is *assumed.* Commitment to a plan is not endogenous in this setup. However insightful it may be, this approach has this as its main weakness.

Notwithstanding this, we will solve system (II). Define

$$W(c_t, n_t, \lambda) = u(c_t, n_t) + \lambda [u_c(c_t, n_t) c_t + u_n(c_t, n_t) n_t].$$

Then we can re-state the problem as

$$\max_{\{c_t, n_t, K_{t+1}\}_{t=0}^{\infty}, \lambda} \left\{ \sum_{t=0}^{\infty} \beta^t W(c_t, n_t, \lambda) - \lambda u_c(c_0, n_0) R_0^K K_0 \right\}$$

$$\text{s.t. } c_t + g_t + K_{t+1} = F(K_t, n_t) + (1-\delta) K_t, \forall t.$$

The term $u_c(c_0, n_0)$ in the objective function is endogenous, where as $R_0^K K_0$ is exogenous. λ is the Lagrange multiplier of the "implementability" constraint in (II).

Taking first order conditions, we should keep in mind that we do not know whether they are sufficient or not, but unfortunately we have no choice but to disregard this problem for the moment. We have

$$\begin{aligned}
c_t &: \beta^t W_c(c_t, n_t, \lambda) = \mu_t, \ t \geq 1 \\
& \quad W_c(c_0, n_0, \lambda) - \lambda u_{cc}(c_0, n_0) R_0^K K_0 = \mu_0 \\
n_t &: \beta^t W_n(c_t, n_t, \lambda) = -\mu_t F_n(K_t, n_t), \ t \geq 1 \\
& \quad W_n(c_0, n_0, \lambda) - \lambda u_{cn}(c_0, n_0) = -\mu_0 F_n(K_0, n_0) \\
K_{t+1} &: \mu_t = [F_k(K_{t+1}, n_{t+1}) + 1 - \delta] \mu_{t+1}, \forall t.
\end{aligned}$$

Observe that for $t = 0$ the first order conditions are different (which reflects the time inconsistency of the choice). Rearranging,

$$-\frac{W_n(c_t, n_t, \lambda)}{W_c(c_t, n_t, \lambda)} = F_n(K_t, n_t), t \geq 1$$

$$W_c(c_t, n_t, \lambda) = \beta W_c(c_{t+1}, n_{t+1}, \lambda) [F_k(K_{t+1}, n_{t+1}) + 1 - \delta], t \geq 1. \tag{13.11}$$

Suppose that the ratio

$$\frac{W_c\left(c_t,\, n_t,\, \lambda\right)}{u_c\left(c_t,\, n_t\right)}$$

is constant over time. Then, equation (13.11) can be rewritten as:

$$u_c\left(c_t,\, n_t\right) = u_c\left(c_{t+1},\, n_{t+1}\right) \beta \left[F_k\left(K_{t+1},\, n_{t+1}\right) + 1 - \delta\right],\ t \geq 1. \tag{13.12}$$

This is the "usual" Euler equation - the one a social planner would choose in an economy without taxes. Compare this expression with equation (13.10). Clearly, (13.10) and (13.12) jointly imply that $\theta_t = 0$ for $t \geq 2$.

What is the intuition for this result? One of the general principles of taxation states that the taxes should be levied on the goods that are less elastically supplied. Clearly, from the perspective of $t = 0$ capital in the distant future is supplied very elastically, since it is relatively easy for consumers to gradually reduce capital stock. In contrast, labor supply cannot be as easily adjusted, since it yields income each period, and such an adjustment would immediately hurt utility. So, to finance a given stream of public expenditures it is preferable to tax labor income and leave capital income untaxed.

The previous argument relies on the hypothesis that the ratio $\frac{W_c(c_t,\, n_t,\, \lambda)}{u_c(c_t,\, n_t)}$ remains constant over time. When will this be valid? There are two answers:

1. Clearly this ratio will not change in the steady state.

2. Some functional forms for the utility representation will also yield such a stationary result. Examples of such functions are:

$$u(c,\, n) = \frac{c^{1-\sigma}}{1-\sigma} + v(n)$$

   or

$$u(c,\, n) = \frac{c^{1-\sigma}}{1-\sigma} \left(1-n\right)^{\gamma(1-\sigma)}.$$

   (The total labor endowment is normalized to 1.)

13.3 References

Atkeson, Andrew, V.V. Chari and Patrick J. Kehoe, "Taxing capital income: a bad idea," *Federal Reserve Bank of Minneapolis Quarterly Review*, 23(1999), 3-17.

Barro, Robert J., "Are government bonds net wealth?" *Journal of Political Economy*, 82(1974), 1095-1117.

Chapter 14

Aggregation

The representative-agent model, which is the focus of much of the above discussion, is very commonly used in macroeconomics. An important issue is whether the inclusion of various forms of consumer heterogeneity leads to a model with similar properties. For example, suppose that consumers have heterogeneous functions $u(c)$, say, all within the class of power functions, thus allowing differences in consumers' degrees of intertemporal substitution. Within the context of the neoclassical model and competitive trading, how would this form of heterogeneity influence the properties of the implied aggregate capital accumulation, say, expressed in terms of the rate of convergence to steady state? This specific question is beyond the scope of the present text, as are most other, similar questions; for answers, one would need to use specific distributional assumptions for the heterogeneity, and the model would need to be characterized numerically. Moreover, the available research does not provide full answers for many of these questions.

For one specific form of heterogeneity, it is possible to provide some results, however: the case where consumers are heterogeneous only in initial (asset) wealth. That is, there are "rich" and "poor" consumers, and the question is thus how the distribution of wealth influences capital accumulation and any other aggregate quantities or prices. We will provide an aggregation theorem which is a rather straightforward extension of known results from microeconomics (Gorman aggregation) to our dynamic macroeconomic context. That is, we will be able to say that, if consumers' preferences are in a certain class, then "wealth heterogeneity does not matter", i.e., aggregates are not influenced by how total wealth is distributed among consumers. Therefore, we can talk about robustness of the representative-agent setting at least in the wealth dimension, at least under the stated assumptions.

14.1 Inelastic labor supply

Consider the following maximization problem:

$$\max_{\{a_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(a_t + w_t - q_t a_{t+1})$$

with a_o given. This problem will be used to represent a consumer's problem in a neoclassical context, but it can also be viewed within a context with no intertemporal production. We will represent the problem recursively and let the aggregate state variable be A (for now unspecified).

The dynamic-programming version is

$$V(a, A) = \max_{a'} u(a + \epsilon w(A) - q(A)a') + \beta V(a', A'),$$

with $A' = G(A)$ for some given G; thus, A is the aggregate variable (possibly a vector) that determines the wage and the bond price.[1] Note that the consumer has ϵ units of labor to supply to the market; we will also discuss the case where consumers differ in their values for ϵ (consumer-workers have different labor productivity).

The task now is to show that with certain restrictions on preferences, this problem has a solution with the feature that individual saving is *linear* in the individual state a, so that the *marginal propensities to save* (and consume) are the same for all consumers provided that they all have the same preferences. Given this, total saving cannot depend on the distribution. The present model does not "explain" where initial asset inequality comes from; it is a primitive of the environment. We will discuss this topic in a later chapter of this text.

The preference class we consider has $u(c) = \hat{u}(A + Bc)$, where A and B are scalars and $\hat{u}$ is (i) exponential, (ii) quadratic, or (iii) CEIS (i.e., $\hat{u}(c) = (1-\sigma)^{-1}(c^{1-\sigma} - 1)$); moreover, we presume interior solutions.

What we need to show, thus, is the following: optimal saving, as summarized by the decision rule $g(a, A)$ to the above recursive problem, satisfies $g(a, A) = \mu(A) + \lambda(A)a$, where μ λ are functions to be determined. Here, thus, $\lambda(A)$ is the marginal propensity to save, and it is equal for agents with different values of a, i.e., for consumers with different wealth levels. We will proceed by a guess-and-verify method; we will stop short of a full proof, but at least provide the key steps.

We make the arguments based on the functional Euler equation, which reads, for all (a, A),

$$q(A)u'(a + \epsilon w(A) - q(A)g(a, A)) = \beta u'(g(a, A) + \epsilon w(G(A)) - q(G(A))g(g(aA), G(A))).$$

For a given G, thus, this equation solves for g.

We will restrict attention to one of the preference specifications; the remaining cases can be dealt with using the same approach. Thus, let $u(c) = (1-\sigma)^{-1}(c^{1-\sigma} - 1)$, so that we can write

$$\left(\frac{q(A)}{\beta}\right)^{\frac{1}{\sigma}} = \frac{a + \epsilon w(A) - q(A)g(a, A)}{g(a, A) + \epsilon w(G(A)) - q(G(A))g(g(aA), G(A))}.$$

[1] Equivalently, one can think of the consumer as choosing "capital" at relative price 1 in terms of consumption units, thus with an ex-post return r which equals $1/q$ or, more precisely, $r(G(A)) = 1/q(A)$.

Using the guess that the decision rule is linear, we see that the functional equation will have a right-hand side which is a ratio of two functions which are affine in a:

$$\left(\frac{q(A)}{\beta}\right)^{\frac{1}{\sigma}} = \frac{B_1(A) + B_2(A)a}{C_1(A) + C_2(A)a},$$

with $B_1(A) = \epsilon w(A) - q(A)\mu(A)$, $B_2(A) = 1 - q(A)\lambda(A)$, $C_1(A) = \mu(A) + \epsilon w(G(A)) - q(G(A))(\mu(G(A)) + \lambda(G(A))\mu(A))$, and $C_2(A) = \lambda(A) - q(G(A))\lambda(G(A))\lambda(A)$. The key now is the following: for this functional equation to be met for all a, we need

$$\frac{B_2(A)}{B_1(A)} = \frac{C_2(A)}{C_1(A)}$$

for all A, and for it to furthermore hold for all values of A, we need

$$\left(\frac{q(A)}{\beta}\right)^{\frac{1}{\sigma}} = \frac{B_2(A)}{C_2(A)}$$

for all A. These two new functional equations determine μ and λ. We will not discuss existence; suffice it to say here that there are two functional equations in two unknown functions. Given this, the key really is that we have demonstrated that the conjectured linearity in a is verified in that the functional Euler equation of the consumer is met for all a under the conjecture.

To obtain some additional insight, we see that the second of the functional equations can be explicitly stated as

$$\left(\frac{q(A)}{\beta}\right)^{\frac{1}{\sigma}} = \frac{1 - q(A)\lambda(A)}{\lambda(A) - q(G(A))\lambda(G(A))\lambda(A)}.$$

We thus see that the marginal propensity function, λ, can be solved for from this equation alone; μ can then be solved recursively from the first of the functional equations.

Several remarks are worth making here. First, ϵ does not appear in the equation for λ. Thus, consumers with different labor productivity but the same preferences have the same marginal propensities to save and consume. Second, σ and β do matter: consumers with different values for these parameters will, in general, have different saving propensities. They will, however, still have constant propensities. Third, suppose that we consider $\sigma = 1$, i.e., logarithmic preferences. Then we see that the functional equation is solved by $\lambda(A) = \beta/q(A)$, i.e., the solution is independent of G and dictates that the marginal savings propensity is above (below) one if the subjective discount rate is lower (higher) than the interest rate. We also see, fourth and finally, that when the consumer is in a stationary environment such that $G(A) = A$, then $\lambda(A) = (\beta/q(A))^{1/\sigma}$. A special case of this, of course, is the "permanent-income" case: when the subjective discount rate equals the interest rate, then any additional initial wealth is saved and only its return is consumed.

Looking at the neoclassical environment, suppose that A is the vector of asset holdings of n different subgroups of consumers within each of which the initial asset holdings are the same, as represented by the values A_i at any arbitrary date. Let ϕ_i be the fraction of consumers of type i. We know, since the economy is closed, that $\sum_{i=1}^{n} \phi_i A_i = K$. Thus, we conjecture that μ and λ depend on K only, and we see that this conjecture is verified: $K' = \sum_{i=1}^{n} \phi_i(\mu(K)+\lambda(K)A_i) = \lambda(K)K+\sum_{i=1}^{n} \phi_i\mu(K)$, with λ and μ solving the functional equations above. This explicitly shows aggregation over wealth: tomorrow's capital stock does not depend on anything but today's capital stock, and not on how it is distributed across consumers. Prices (q and w), of course, since they are given by marginal products of aggregate production, also depend only on K in this case, which is why μ and λ will only depend on K.

14.2 Valued leisure

Does aggregation obtain when preferences allow leisure to be valued, so that potentially different consumers supply different amounts of leisure? In this case, aggregation also requires that the total amount of labor supplied depend on K only, and not on the wealth distribution. As in the case of total saving, this will occur if consumers' individual labor supplies are linear in their individual asset (wealth) levels.

We will not provide full proofs; these would proceed along the lines of the above arguments. There are two separate cases to look at. In one, there are wealth effects on labor supply; in the other, there are no wealth effects. Both seem relevant, since it is not clear how large such effects are.

14.2.1 Wealth effects on labor supply

Suppose that period utility satisfies

$$u(c, l) = \hat{u}(A + g(c - \bar{c}, l - \bar{l}))$$

where $\hat{u}$ is in the class above (exponential, quadratic, or with CEIS), g is homogeneous of degree one in both arguments, and A, $\bar{c}$, and $\bar{l}$ are scalars. Then it is possible to show that aggregation obtains. The reason why this preference formulation leads to aggregation is that the first-order condition for the leisure choice will deliver

$$\frac{g_2(1, z)}{g_1(1, z)} = w(A)\epsilon$$

where $z \equiv (l - \bar{l})/(c - \bar{c})$; thus, all consumers with the same preferences and ϵs will have the same value for z at any point in time. This means that there is aggregation: total labor supply will be linear in total consumption. Formally, we let consumers first maximize over the leisure variable and then use a "reduced form" $g(c - \bar{c}, l - \bar{l}) = (c - \bar{c})g(1, z(A))$, which is of a similar form to that analyzed in the previous section.

The functional form used above allows us to match any estimated labor-supply elasticity; the use of a g which is homogeneous of degree one is not too restrictive. For example, a CES function would work, i.e., one where $\rho \log g(x, y) = \log\left(\varphi x^\rho + (1 - \varphi)x^\rho\right)$.

14.2.2 Wealth effects on labor supply

Now consider a case where $u(c,l) = \hat{u}(A+Bc+v(l))$. Here, the first-order condition delivers

$$v'(l) = Bw(A)\epsilon.$$

The key here, thus, is that all consumers (with the same v and the same ϵ) choose the same amount of leisure, and therefore the same amount of hours worked. Again, we obtain a reduced form expressed in terms of individual consumption of the type above.

Alternatively, consider $u(c,l) = \hat{u}\left(c^{\alpha_c} + Bl^{\alpha_l}\right)$. If $\alpha_c = \alpha_l$ we are in the first subclass considered; if $\alpha_c = 1$, we are in the second.[2] With $\alpha_c \neq \alpha_l$ and both coefficients strictly between zero and one, we do not obtain aggregation.

[2]The case $\alpha_l = 1$ also delivers aggregation, assuming interior solutions, but this case has the unrealistic feature that c does not depend on wealth.